# تقنيات اللغة العربية الحاسوبية

## معايير التقييم ورؤى التطوير

دراسات لغوية حاسوبية في

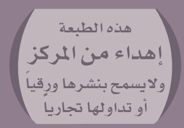
المدققات الإملائية والنحوية
محركات البحث
آليات التعرف الآلي على الحروف
آليات الإحصاء اللغوية



تأليف: د. عمرو جمعة

الدراسات ٤




www.kaica.org.sa

الدراسات -٤-

# تقنيات اللغة العربية الحاسوبية

## - معايير التقييم ورؤى التطوير -

**دراسات لغوية حاسوبية في**

(المدققات الإملائية والنحوية - محركات البحث - آليات التعرف الآلي على الحروف - آليات الإحصاء اللغوية)

تأليف

**د. عمرو جمعة**

دكتوراه اللغويات الحاسوبية






# تقنيات اللغة العربية الحاسوبية

## - معايير التقييم ورؤى التطوير -

الطبعة الأولى
1437 هـ - 2016م



المملكة العربية السعودية - الرياض
ص.ب ١٢٥٠٠ الرياض ١١٤٧٣
هاتف:٠٠٩٦٦١١٢٥٨٧٢٦٨ - ٠٠٩٦٦١١٢٥٨١٠٨٢
البريد الإليكتروني: nashr@kaica.org.sa

ح/ مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، ١٤٣٧هـ.
فهرسة مكتبة الملك فهد الوطنية أثناء النشر
جمعة، عمرو
تقنيات اللغة العربية الحاسوبية./ عمرو جمعة- الرياض، ١٤٣٧هـ
..ص؛ .. سم
ردمك: ٥-٦-٩٠٧٢٠-٦٠٣-٩٧٨
١- الحواسيب ٢- استخدام اللغة العربية أ. العنوان
ديوي ٠٠٦,٣٥ ١٤٣٧/٤٦٧٣
رقم الإيداع: ١٤٣٧/٤٦٧٣
ردمك: ٥-٦-٩٠٧٢٠-٦٠٣-٩٧٨

التصميم والإخراج
دار وجوه للنشر والتوزيع
Wojooh Publishing & Distribution House
www.wojoooh.com
المملكة العربية السعودية - الرياض
الهاتف:4562410 الفاكس:4561675
للتواصل والنشر:
info@wojoooh.com



# مقدمة الكتاب

بسم الله والحمد لله والصلاة والسلام على سيدنا رسول الله، وبعد فهذا هو الكتاب الثاني في سلسلة كتب المؤلف في معالجة اللغة العربية آليا، وهو بعنوان «تقنيات اللغة العربية الحاسوبية – معايير التقييم ورؤى التطوير»، يعرض فيه المؤلف لمعايير تقييم بعض التقنيات اللغوية الحاسوبية مع تقديم رؤى تطويرية لبعضها، وآفاق جديدة لاستخدام بعضها. والكتاب كما يبدو من عنوانه يدور في فلك معالجة اللغات الطبيعية؛ متخذا من اللغة العربية ومن تقنياتها منطلقًا جديدًا لآفاق بحثية جديدة.

ويهدف هذا الكتاب إلى تقديم منهج لغوي لتقييم التقنيات اللغوية الحاسوبية التي تعالج اللغة العربية، من خلال عدد من المعايير اللغوية، فنظرة واحدة على بعض التقنيات العربية كافية لإدراك الهوة الواسعة بين مخرجات هذه التقنيات وبين ما يجب أن تكون عليه النتائج المرجوة، ومن ثم دعت الحاجة إلى وضع لبنة لغوية لمعايير تقييم هذه التقنيات، وستقتصر هذه المعايير على الجانب اللغوي فقط تاركة الجانب الفني للمتخصصين فيه، فعند الحديث عن المعايير اللغوية لتقييم محركات البحث العربية على الشبكة العالمية – على سبيل المثال – سيقتصر تقييم محركات البحث على المعايير التي ينبغي توافرها في أي محرك بحثي عربي عند التعامل مع كلمات البحث العربية ونتائجها، وهو ما سيجعل نتائج البحث صحيحة وسليمة.

كما يهدف الكتاب إلى تقديم رؤى تطويرية جديدة لتطوير بعض التقنيات اللغوية العربية، يمكن من خلالها رفع كفاءة هذه التقنيات وتحسين جودة مخرجاتها، كما هو الحال مع تطبيقات التعرف الآلي على الحروف العربية، أو يمكن من خلالها إبراز الدور التعليمي لهذه التقنيات الحاسوبية، كما هو الحال مع المدققات الإملائية والنحوية؛ إذ يمتد دور هذه التقنيات اللغوية العربية إلى الإسهام في تعليم مستخدميها من العرب قواعد العربية كذلك، إضافةً إلى دورها الأساسي في معالجة العربية آليًّا.

ويعرض الكتاب لآفاق جديدة يمكن استخدام بعض التقنيات اللغوية فيها؛ لتحقيق الاستفادة القصوى من هذه التقنيات اللغوية وربطها بواقع المستخدم العربي، فمن هذه الآفاق - على سبيل المثال - مراجعة النص القرآني وتدقيقه آليًّا، ومنها استخدام تقنيات الإحصاء اللغوية في تحليل مدونات النصوص العربية الضخمة والاستفادة من نتائج هذا التحليل في صياغة مناهج تعليمية للعربية المعاصرة، تتخذ من واقع اللغة أساسًا لتعليمها.

ويقع هذا الكتاب في مقدمة وثلاثة فصول؛ تتناول المقدمة التعريف بالكتاب وموضوعاته ومادته العملية وتقسيم فصوله ومباحثه، وتعرض لأهداف الكتاب والغرض من تأليفه وللتقنيات التي اعتمد عليها.

أما الفصل الأول فهو بعنوان «معايير تقييم بعض التقنيات اللغوية» ويتناول المعايير اللغوية لتقييم ثلاثة من التقنيات اللغوية، ويقع في ثلاثة مباحث؛ المبحث الأول: المعايير اللغوية لتقييم برامج التدقيق الإملائي والنحوي، والمبحث الثاني: المعايير اللغوية لتقييم محركات البحث العربية على شبكة الإنترنت، والمبحث الثالث: المعايير اللغوية والفنية لتقييم برامج التعرف الضوئي على الحروف العربية.

وعنوان الفصل الثاني: «رؤى مطروحة لتطوير بعض التقنيات اللغوية» ويتناول رؤيتين لتطوير تقنيتين من التقنيات اللغوية الحاسوبية، ويقع في مبحثين اثنين؛ المبحث الأول يعرض مقترحا لتطوير تقنيات التدقيق الإملائي والنحوي لتفعيل دورها في تعليم العربية. والمبحث الثاني يقدّم مقترحا لتطوير آليات التعرف الآلي على الحروف العربية.

والفصل الثالث بعنوان: «آفاق استخدام بعض التقنيات اللغوية» ويتناول أفقين تطبيقيين من آفاق استخدام التقنيات اللغوية، ويقع في مبحثين اثنين؛ المبحث الأول في

مراجعة النص القرآني وتدقيقه آليًّا في دراسة لغوية حاسوبية، والمبحث الثاني في تحليل النصوص باستخدام تقنيات الإحصاء اللغوية.

وفي الخاتمة يعرض المؤلف أهم النتائج التي توصل إليها بحثه، وأهم التوصيات التي يوصي بها الدارسين والباحثين في مجال معالجة اللغة العربية.



# الفصل الأول
# معايير تقييم التقنيات اللغوية

## المبحث الأول
## المعايير اللغوية لتقييم برامج التدقيق الإملائي والنحوي

### تمهيد:

المدققان الإملائي والنحوي آليتان برمجيتان تتعاملان مع النصوص ( في الكلمة والتركيب معًا)، وتتلخص وظيفتاهما في مراجعة صحة الكلمة إملاءً وإعرابًا وصحة الجملة تركيبًا، وفي اقتراح بدائل صحيحة للكلمة غير الصحيحة إملاءً أو نحوًا، أو اقتراح تعديل للجملة غير الصحيحة تركيبًا؛ لأجل الوصول إلى الصواب النحوي أو التركيبي. ويمكن إجمال المعايير اللغوية لتقييم برامج التدقيق الإملائي والنحوي فيما يأتي:

### المعايير اللغوية لتقييم برامج التدقيق الإملائي والنحوي:

**١- مدى توافر عدة مستويات لعمل المدقق الإملائي:**

إذ ينبغي أن تتحقق جليًّا في تطبيقات التدقيق الإملائي مستويات التدقيق، ومن الممكن اقتراح ثلاثة مستويات - قابلة للزيادة - على النحو الآتي:

١- التقيّد بقواعد الإملاء، ويعني: التقيّد بهمزة الألف الابتدائية أو بنقطتي الياء النهائية - على سبيل المثال -.

٢- المرونة في التقيّد بقواعد الإملاء، ويعني التقيّد بهمزة الألف الابتدائية في

خيارات التدقيق الإملائي (في حالة الكتابة المصرية) - على سبيل المثال-، وكذلك التقيد بنقطتي الياء النهائية في خيارات التدقيق الإملائي (في حالة الكتابة الشامية) - على سبيل المثال-.

٣- مراعاة الأخطاء الشائعة، ويعني عدم التقيد بهمزة الألف الابتدائية أو بنقطتي الياء النهائية- على سبيل المثال-.

**٢- مدى معالجة تطبيقات التدقيق الإملائي لأنماط الأخطاء الإملائية الشائعة في النصوص العربية ومن أهمها:**

- الخلط بين أشكال همزتي القطع والوصل (أ،إ،ا،آ)، مثل كتابة كلمة «استقبال» بهمزة الوصل هكذا: «إستقبال»، «أستقبال»، «آستقبال».

- الخلط بين الألف المقصورة (ى) والياء المنقوصة (ي)، مثل كتابة كلمة «مرتضى» (بياء غير منقوطة) هكذا: «مرتضي» (بنقطتين).

- الخطأ في رسم الهمزات المتوسطة والمتطرفة (ء، ئ، ؤ، أ، آ)، مثل كتابة كلمة «شيء» بهمزة على السطر هكذا: «شيئ» بهمزة على نبرة.

- الخلط بين التاء المربوطة (ـة) والهاء (ـه)، مثل كتابة كلمة «رماية»؛ هكذا: «رمايه» بهاء وليست بتاء مربوطة.

- اتصال كلمتين دون مسافة بينهما، مثل كتابة كلمتي «اتصال كلمتين» هكذا: «اتصالكلمتين».

- وجود نقص في حروف الكلمة، مثل كتابة كلمة «المحامي» هكذا: «المحام».

- وجود تكرار أو زيادة في الكلمة، مثل كتابة كلمة «مشكور» هكذا: «مشكووور».

- وجود قلب بين حروف الكلمة الواحدة، مثل كتابة كلمة «مشكور» هكذا: «مكشور».

- الحروف التي تنطق لهجيًّا بطريقة مخالفة لكتابتها، فحرف «الضاد» في بعض البلاد العربية ينطقها البعض «ظاء» ومن ثم يكتبونها «ظاء»، ومما رأيته على مواقع التواصل الاجتماعي عنوانا للكتاب المشهور «في ضلال القرآن»؛ بالضاد لا بالظاء؛ تأثرا باللهجة والصواب «في ظلال القرآن».

- الخلط بين الحروف المتجاورة في لوحة مفاتيح الحاسوب، حيث يخطئ

البعض في كتابة بعض الحروف نتيجة لسرعته في الكتابة على لوحة المفاتيح، مثل كتابة كلمة «فاروق» هكذا: «فاروف» وذلك لتجاور حرفي «الفاء» و»القاف» على لوحة المفاتيح، ويخطئون كذلك في الحروف التي تكتب بطريقة مركبة على لوحة المفاتيح مثل كتابة كلمة «إيمان» هكذا: «غيمان» حيث تكتب الهمزة المكسورة عن طريق الضغط على «غshift + "، وفي حال عدم الضغط على مفتاح "shift" تكتب "غينًا" على الفور.

**٣- مدى معالجة تطبيقات التدقيق الإملائي للبسين الصرفي والدلالي وحل مشكلاتهما:**

فكثير من مستخدمي تطبيقات التدقيق الإملائي يضيقون ذرعًا بتطبيقات التدقيق الإملائي، وذلك بسبب فشلها في اكتشاف الكثير من الأخطاء الإملائية لبعض الكلمات في سياقات معينة، واعتبارها كلمات صحيحة على الرغم من كونها خطأً في هذه السياقات.

فكلمة «اضرب» بهمزة الوصل في سياق جملة «دعوني اضرب لكم مثلا»؛ تعد خطأ إملائيًّا يلزم تصويبه إلى الكلمة «أضرب» بهمزة القطع، ومع ذلك تقف تطبيقات التدقيق الإملائي عاجزة عن تصويب هذه الكلمة. وكذلك يتجاهل المدقق الخطأ الإملائي في كلمة «السنة» في سياق التركيب «السنة النيران»، وهو ما يعني الفشل في التعرف على الخطأ واقتراح بدائل صحيحة لتصويبه.

ويرجع السبب في ذلك إلى اللبس الصرفي، فالمدقق الإملائي يتعامل مع الكلمة المفردة على مستواها الصرفي فقط، ومن ثم فإن أية احتمالات صرفية للكلمة المراد تدقيقها صحيحة لغويًّا، بلا مراعاة للسياق اللغوي للكلمة، وهو ما يحمل المدقق على عدم الحكم بخطأ الكلمتين «اضرب» و»السنة» إملائيا، على الرغم من خطأ استخدامهما في هذين السياقين.

إن التباس الشكل الكتابي لمضارع الثلاثي المكتوب خطأً هكذا (اضرب) مع كلمة أخرى صحيحة الشكل هي (اضرب – أمر الثلاثي) قد تسبب في تجاهل المدقق الإملائي للخطأ الإملائي ظاهرا (السياقي حقيقة) وعدم اكتشافه ومن ثم اقتراح البدائل الصحيحة له.

أما هذا الالتباس فيرجع إلى عدم اعتماد المدقق الإملائي على السياق اللغوي والدلالي عند تحليله للخطأ الإملائي واعتماده فقط على المحلل الصرفي الذي يتعامل مع الكلمة دون النظر إلى سياقها. ففي جملة «شعب من فصيلة الصبار» -على سبيل المثال- لا يمكن للمدقق أن يكتشف الخطأ في كلمة «شعب» ومن ثم لا يقترح تصويبها إلى «عشب»، وذلك لأن كلمة «شعب» صحيحة صرفيا، ولكنها خطأ مع هذه المصاحبة، وليس من سمات المدققات الإملائية الحالية أن تعتمد على المصاحبات المعجمية بل تكتفي فقط بالتحليل الصرفي للكلمة دون سياقها الدلالي.

ومن أمثلة تلك التراكيب؛ «بنات نادر» وصوابها «نبات نادر»، و«تعقدت الجرحى»، وصوابها «تفقدت الجرحى»، فالكلمتان «بنات»، و«تعقدت» كلمتان صحيحتان صرفيا، وهو ما حمل المدقق على الحكم بصحتهما، لكن السياق وحده هو ما يحكم بخطأ الكلمتين.

**٤- مدى معالجة تطبيقات التدقيق النحوي لآثار اللبس التركيبي:**

رأينا كيف يقف اللبس عقبة كأداء دون تحقيق الغاية من بعض التطبيقات اللغوية الحاسوبية العربية، وليست تطبيقات التدقيق النحوي بعيدة عن ذلك؛ إذ يؤثر اللبس التركيبي على قيام تطبيقات التدقيق النحوي هذه بعملها. فالمدقق الإملائي والنحوي لا يستطيع - على سبيل المثال - تصويب علامة إعراب كلمة «بالوَلَدان» (مثنى ولد) إلى «بالولدين»، وذلك لالتباس كلمة «بالوَلَدان» بكلمة «بالوِلْدان» جمع التكسير من «ولد»، بل يتعدى الأمر ذلك إلى علاقاتها التركيبية في الجملة، فصفة كلمة «بالوَلَدان» (مثنى ولد) هي كلمة مثنى مذكر، تقول: «بالوَلَدان المُخَلَّدَينِ» أما صفة كلمة «بالوِلْدان» فهي جمع مذكر أو جمع تكسير، تقول: «بالوِلْدان المُخَلَّدِينَ».

وثنائيات الكلمات (لهب، لهب / وعيد، وعيد / بطاقة، بطاقة / فطائر، فطائر) لا يستطيع المدقق النحوي العرف على علاقاتها التركيبية لاحتمالاتها الصرفية ومن ثم التركيبية المتعددة ،

فكلمة «لهب» تحتمل أن تكون من باب:

لهب: لَهَبٌ، اسم من معانيه: ما يرتفع من النار كأنه لسان[١].

أو تكون من باب:

لهب: لَهَبَّ: لام التعليل + هَبَّ (فعل ماض، من معانيه: أسرع ونشط، ويقال هب فلان إلى الشيء: نهض إليه)[٢].

وكلمة «وعيد» تحتمل أن تكون من باب:

وعيد: وعيد، مصدر الفعل وَعَدَ، يقال وعد فلانا الشر، ووعده به وعيدا، بمعنى: هدده به[٣].

أو تكون من باب:

وعيد: الواو حرف عطف + عيد (كل يوم يحتفل فيه بذكرى كريمة أو حبيبة)[٤].

وكلمة «بطاقة» تحتمل أن تكون من باب:

بطاقة: (رقعة صغيرة من الورق وغيره، يكتب عليها اسم الشخص وعنوانه أو بيان ما تعلق عليه[٥].

أو تكون من باب:

بطاقة: الباء حرف عطف + طاقة (الطاقة القدرة، وما يستطيع الإنسان أن يفعله بمشقة)[٦].

وكلمة «فطائر» تحتمل أن تكون من باب:

فطائر: (جمع فطيرة)، والفطيرة خبز تؤدم بزبد أو نحوه[٧].

---

١ – المعجم الوسيط، إبراهيم مصطفى، أحمد حسن الزيات، حامد عبدالقادر، محمد علي النجار، مجمع اللغة العربية بالقاهرة، ط.٤، ٢٠٠٤، ص ٨٧٤، مادة «لهب».

٢ – نفسه، ص ١٠٠٧، مادة «هبب».

٣ – نفسه، ص ١٠٨٥، مادة «وعد».

٤ – نفسه، ص ٦٥٨، مادة «عود».

٥ – نفسه، ص ٦٣، مادة «بطق».

٦ – نفسه، ص ٥٩٢، مادة «طوق».

٧ – نفسه، ص ٧٢٠، مادة «فطر».

**أو تكون من باب:**

فطائر: الفاء حرف عطف + طائر (وهو من الحيوان كل ما يطير في الهواء بجناحين)[1].

ومن ثم وجب على تطبيقات التدقيق الإملائي والنحوي معالجة اللبس اللغوي بأنواعه المختلفة الصرفية والتركيبية والدلالية، واعتبار ذلك معيارًا مهمًّا من معايير تقييمها.

**٥- مدى نجاح عملية التصحيح التلقائي على المستوى الإملائي:**

يطرح المدقق الإملائي بدائل عديدة مقترحة تصحيحًا للخطأ الإملائي المكتشف في النص؛ من خلال نافذة صغيرة تحوي كلمات أخرى مقترحة لتصحيح الخطأ الموجود في حالة الخطأ الإملائي، وعادة ما ترتب هذه الاقتراحات ترتيبًا تنازليًّا بدءًا بأوسعها انتشارًا وانتهاءً بالأقل. ويعتمد التصحيح التلقائي على ترتيب هذه البدائل ترتيبًا صحيحًا؛ إذ يختار المصحح التلقائي عند تفعيله من قبل المستخدم البديل الأول؛ خيارًا وحيدا للتصحيح في كل المواضع، ومن ثم فإن نجاح المصحح التلقائي في أداء مهمّته يعتمد في الأساس على الدقة في ترتيب البدائل. ولذا يمكن اعتبار مدى نجاح المدقق الإملائي في التصحيح التلقائي من معايير تقييم هذا المدقق.

**٦ - مدى تعلم المدقق الإملائي والنحوي من تصحيحات المستخدم ومدى استفادته من هذه التصحيحات في التصحيح التلقائي:**

فهناك مستوى تُزوَّد فيه بعض المدققات الإملائية بملفات تعمل بوصفها قاعدة بيانات معجمية لكلمات جديدة يحكم المدقق بخطئها على الرغم من صحتها لدى المستخدم، ومن ثم يمكِّن المدققُ المستخدمَ من إضافة هذه الكلمات إلى قاعدة بيانات خاصة بهذه النسخة من البرنامج، يمكن من خلالها للمدقق الحكم بصحة هذه الكلمات عند تكرارها.

وثمة مستوى آخر يمكن عدّه معيارًا من معايير تقييم المدقق الإملائي والنحوي وهو مدى الاستفادة من تصحيحات المستخدم في التصحيح التلقائي، إذ يزداد المدقق

---

١ - نفسه، ص ٥٩٥، مادة «طير».

خبرة كلما استُعمل مرة بعد مرة، فهو يخزّن خبراته لكي يستعملها مستقبلًا. فإذا ما قام المستخدم بتعديل خطأ معين سبق أن اكتشفه المدقق، فعلى المدقق حينئذ تصحيح الخطأ نفسه عند تكراره وفقا لتصحيح المستخدم.

**٧ - مدى جاهزية المدقق الإملائي والنحوي للتعامل مع علامات الضبط بالشكل بنيةً وإعرابًا:**

وهو معيار قلّما يتوافر في المدققات الإملائية والنحوية، وربما كان ذلك بسبب تخلي العربية المعاصرة عن علامات الضبط بالشكل، ولكنه ضروري في معالجة النصوص التراثية التي تحافظ على علامات الضبط بالشكل.



## المبحث الثاني

## المعايير اللغوية لتقييم محركات البحث العربية على شبكة الإنترنت

### تمهيد:

تعد نظم البحث واسترجاع المعلومات أحد أهم الطرق التي نلجأ إليها للتغلب على مشكلة فوضى المعلومات أو تضخم الإنتاج الفكري على الشبكة العالمية. ويمكن لمستخدمي محركات البحث التي تتعامل مع العربية على شبكة الإنترنت؛ أن يدرك الهوة الواسعة بين مخرجات هذه المحركات وبين ما يجب أن تكون عليه النتائج المرجوة فيما يخص البحث باللغة العربية.

وسيركز هذا المبحث على المعايير اللغوية لتقييم محركات البحث العربية على الشبكة العالمية، وسيقتصر تقييم محركات البحث لغويًّا - في هذا المبحث - على المعايير التي ينبغي توافرها في أي محرك بحثي عربي عند التعامل مع كلمات البحث دون التطرق إلى نتائج البحث، ففي حالة نجاح المحرك في التعامل مع كلمات البحث وفق معايير لغوية محددة، فإن من شأن ذلك أن يسفر عن نتائج جيدة وسليمة لكلمات البحث هذه.

ويدرس المبحث تقييم طريقة تعامل محرك البحث مع كلمات البحث على مستويين اثنين؛ مستوى صرفي وآخر دلالي.

**فمن المعايير اللغوية لتعامل محرك البحث مع كلمات البحث على المستوى الصرفي:**

**١ - مراعاة التدقيق اللغوي للمدخلات(Input).**

**٢- مراعاة الأشكال الكتابية:**

مراعاة أشكال الكلمات المعربة.

مراعاة الأشكال اللهجية للكلمات العربية.

ج- مراعاة أشكال لغة الفرانكو آراب.

د- مراعاة أشكال الكلمات المركبة من كلمتين بينهما مسافة.

هـ - مراعاة أشكال الكلمات التي تحوي حروفًا تنطق ولا تكتب.

**٣- مراعاة زوائد الكلمة العربية (من السوابق واللواحق) (Affixes Prefixes- suffixs).**

ومن المعايير اللغوية لتعامل محرك البحث مع كلمات البحث على المستوى الدلالي:

مراعاة الترادف (Synonymy).

مراعاة المشترك اللفظي (Homonymy).

مراعاة اللبس الصرفي على مستوى كلمات البحث وعلى مستوى كلمات التوقف وذلك على التفصيل الآتي:

**المعايير اللغوية لتعامل محرك البحث مع كلمات البحث على المستوى الصرفي:**

ونعني بها المعايير المتعلقة بتعامل محرك البحث مع بنية كلمة البحث الصرفية من حيث صحتها لغويًّا وأشكالها الكتابية المختلفة.

**١- مراعاة التدقيق اللغوي للمدخلات(Input):**

ينبغي لمحرك البحث أن يدقق كلمات المستخدم لغويًّا، تحسبًا وتداركًا لخطأ المستخدم، ومن ثم ينبّه المستخدم إلى الخطأ ويقترح الكلمة الصحيحة لكلمة البحث الخطأ، ويقترح عرض النتائج الخاصة بالكلمة الصحيحة. وهو ما نلحظه في محرك البحث الشهير جوجل.

لكنه – وإن اقترح الصورة الصحيحة للكلمة – لا يهمل الصور الخطأ كتابيًّا، فإن

حصر المحرّك بحثه على الكلمات الصحيحة فقط يقلّل من نتائج البحث ويتسبب في فقد الكثير من النتائج، فليس كل مستخدم عربي على علم بقواعد الكتابة والإملاء اللغوية. وهو ما ينقلنا إلى ضرورة مراعاة الأخطاء الإملائية الشائعة (The common spelling errors)، وسنركز هنا على الأخطاء الإملائية المتعلقة بحروف بعينها، هي:

أ - ا - إ - آ / ء - ؤ - ئ / ي - ى / ة - ه

فكل مجموعة من مجموعات الحروف هذه تتعاور فيما بينها على الكلمة الواحدة - وإن خطأً -، فقد يكتب بعض مستخدمي الإنترنت كلمة «الاسلام» هكذا (همزة وصل)، وقد يكتبها أحدهم هكذا «الأسلام» بهمزة مفتوحة، وقد يكتبها أحدهم صحيحة هكذا «الإسلام» بهمزة قطع مكسورة. وعلى محرك البحث العربي أن يراعي صور كتابة كلمة البحث المختلفة صحيحة كانت أو خطأً.

فإن أي ألف في أول الكلمة (ا) يمكن أن تكتب على ثلاثة أشكال: إ = ا = أ = آ، ومثاله: أيمن - ايمن - إيمن - آيمن.

وإن أي همزة تليها ألف مد (ءا) في أول الكلمة وفي منتصفها يمكن أن تكتب: آ، ومثاله: ءامنوا = آمنوا ، قرءان = قرآن.

وإن أي همزة على السطر تليها واو مد (ءو) في منتصف الكلمة يمكن أن تكتب هكذا: ؤو، ومثاله: عبد الرءوف = عبد الرؤوف.

وإن أي همزة على نبرة تليها واو مد (ئو) في منتصف الكلمة، يمكن كتابتها هكذا: ؤو، ومثاله: شئون = شؤون ، مسئولية = مسؤولية.

وإن أي همزة على نبرة (ئ) في آخر الكلمة يمكن أن تكتب هكذا: يء وهكذا: ىء، ومثاله: شيء = شئ = شيئ = شىء.

وإن أي ياء منقوطة (ي) في آخر الكلمة يمكن أن تكتب غير منقوطة هكذا (ى)، ومثاله: قاضي = قاضى = قاضٍ.

وإن أي تاء مربوطة (ة) في آخر الكلمة يمكن أن تكتب هاء (ه)، ومثاله: مدرسة - مدرسه.

والخلاصة أن من الضروري لأي محرك بحث عربي أن يراعي الأخطاء الإملائية عند البحث بكلمات تحتمل حروفها أشكالًا مختلفة خطأً كانت أو صحيحة، وذلك بأن

يتضمن مثلا بعض القواعد اللغوية التي من شأنها توليد الأشكال اللغوية المختلفة لكلمة البحث؛ الصحيحة منها والخطأ لضمان توفير أكبر قدر من نتائج البحث.

### ٢- مراعاة الأشكال الكتابية:

#### أ- مراعاة أشكال الكلمات المعربة:

تتفاوت صور كتابة بعض الكلمات العربية من بلد إلى آخر، لا سيما الكلمات المعربة، وهو أمر ينبغي مراعاته في نظم البحث والاسترجاع، ومعيار لغوي مهم ينبغي تقييم محرك البحث العربي من خلاله، ومن صور ذلك على سبيل المثال:

بعض الكلمات المعربة عن الإنجليزية، حيث يمكن تعريب الحرف (G) إلى (ج) أو إلى (غ).

فكلمة: Bibliography تكتب «ببليوجرافيا» ـ «ببليوغرافيا»

وكلمة: Google تكتب جوجل ـ غوغل.

وكلمة: kilogram تكتب كيلو جرام – كيلو غرام.

وكلمة: Romantic تكتب رومانتيكية ـ رومانسية ـ رومانطيقية

وكلمة: Michael تكتب ميشيل – ميكائيل – مايكل – ميخائيل.

#### ب- مراعاة الأشكال اللهجية للكلمات العربية:

ولا يجب أن تقتصر معالجة محرك البحث العربي على مراعاة الأشكال الكتابية للكلمات المعربة فحسب بل يجب أن تمتد كذلك إلى أشكال هذه الكلمات في اللهجات العربية المختلفة، فكلمة: «جوجل» يكتبها البعض «قوقل» وكلمة: «رجال» يكتبها آخرون «ريال».

#### ج- مراعاة أشكال لغة الفرانكو آراب:

ينبغي أن ننبه إلى صورة أخرى من صور اختلاف الأشكال الكتابية وهي كتابة الحروف العربية بحروف لاتينية وهو ما يعرف بلغة الفرانكو آراب، وهي لغة يعمد إليها الكثير من الشباب في غرف الدردشة على الإنترنت لعدم تمكنهم من الكتابة بحروف عربية، فيعمدون إلى كتابة حروف العربية بحروف لاتينية مع استبدال بعض الحروف الإنجليزية بأرقام إنجليزية تتشابه وصور الحروف العربية في الشكل، مثل:

ء = ٢ ، ع = ٣ ، غ = '٣ ، خ = ٥ ، ط = ٦ ، ح = ٧ ، خ = '٧ . وهي لغة تخلط بين الحروف العربية والحروف والأرقام الإنجليزية، فكلمة "حوار" على سبيل المثال تكتب هكذا: ٧ewar .

وعلى محرك البحث العربي أن يجعل في الحسبان نقل كلمات البحث العربية إلى هذه اللغة، والبحث عن نتائجها، ونقل كلمات البحث المكتوبة بهذه اللغة إلى العربية والبحث عن نتائجها.

**د- مراعاة أشكال الكلمات المركبة من كلمتين بينهما مسافة:**

في العربية كلمات بحث مركبة من كلمتين بينهما مسافة، مثل "عبد الله" ، "أبو ظبي" . وينبغي على محرك البحث أن يراعي أن من صور كتابة كلمات البحث هذه حذف المسافة جزأي المركب، عبد الله = عبدالله، أبو ظبي = أبوظبي.

وقاعدة ذلك تتلخص في أن أي كلمة من هذه الكلمات المركبة من جزأين يمكن أن تلتصق بما بعدها دون مسافة إذا كان آخر حرف في الكلمة الأولى أحد الحروف التالية: د، ذ، ر، ز، و. وتلخصها القاعدة التالية: أي [كلمة بحث مركبة] = أي [كلمة ١] + مسافة + [كلمة ٢] = أي [كلمة ١ تنتهي بأحد حروف(د، ذ، ر، ز، و)] + مسافة + [كلمة ٢] = أي [كلمة ١ تنتهي بأحد حروف(د، ذ، ر، ز، و)] + أي [كلمة ٢] (بحذف المسافة قبلها) .

**هـ - مراعاة أشكال الكلمات التي تحوي حروفًا تنطق ولا تكتب ومثالها: (لكن = لاكن، داود = داوود، الرحمن = الرحمان) ، فالبحث عن أي من صور الكلمات لابد أن يقتضي البحث عن صورها الأخرى؛ مراعاة لكثير من المستخدمين الذين لا يراعون قواعد الرسم الإملائي.**

### ٣- مراعاة زوائد الكلمة العربية (من السوابق واللواحق)

### :Affixes (Prefixes- suffixs)

فالكلمة في العربية قد تكون مجردة من السوابق واللواحق، وقد تلحقها السوابق واللواحق (كل على حدة أو كلاهما معا)، فكلمة (إيمان) جذع يقبل العديد من السوابق واللواحق، ومن صوره على سبيل المثال: الإيمان - لإيمان - للإيمان - فالإيمان -

والإيمان- لإيمانه - لإيمانهم - لإيمانك - لإيمانكم...

ومن السوابق حروف الجر المتصلة مثل الباء واللام وحروف العطف المتصلة مثل الواو والفاء، ومن اللواحق الضمائر المتصلة (ـه، ـهم ، ـها).

وهذا المعيار يرتبط أيضا بضرورة مراعاة حالات إعراب الكلمة العربية المعربة بالحروف لا بالحركات، فالبحث عن كلمة مثنى ينبغي أن يتضمن البحث عن علامتي التثنية (ان) و (ين)، أي أن البحث عن «المسلمين» = البحث عن «المسلمان»، والبحث عن جمع المذكر السالم يتضمن البحث عن حالتيه الإعرابيتين (ين) ، (ون) ، أي أن البحث عن «المسلمين» = البحث عن «المسلمون».

وكذا الأمر بالنسبة لألف التنوين بالفتحة فالبحث عن «مسلم» = البحث عن «مسلمًا» ، والبحث عن «إيمان» = البحث عن «إيمانًا».

ويمكن لمحرك البحث في هذه الحالة أن يقدم للمستخدم إمكانية الاستعاضة عن الحروف المتغيرة بالكلمة بالعلامة (*) ، على النحو التالي:

فالبحث عن «المؤمن*ن» = البحث عن «المؤمنان» - «المؤمنين» - «المؤمنون»

## المعايير اللغوية لتعامل محرك البحث مع كلمات البحث على المستوى الدلالي:

### ١- مراعاة الترادف (Synonymy):

يعد معيار مراعاة الترادف أهم معيار دلالي من المعايير اللغوية لتقييم محركات البحث العربية، فمستخدم محركات البحث على الإنترنت لا يستخدم الصور المترادفة للكلمة المراد البحث عنها، وهو ما يؤدي إلى فقد كثير من النتائج المفيدة له، بسبب عدم استخدامه لمترادفات كلمته. فالبحث عن كلمة ما في محركات البحث لا يتضمن البحث عن مرادفاتها.

ونقصد بالترادف هنا الكلمات المختلفة التي تحمل دلالات واحدة، وذلك على مستوى الفصحى واللهجات، أو حتى الكلمات المعربة، ومنه على سبيل المثال لا الحصر:

- البترول ، النفط - حقول البترول ، آبار البترول - حقول النفط - آبار النفط.
- قوانين حمورابي ، تشريعات حمورابي ، شريعة حمورابي، تشريع حمورابي.
- بنوك ، مصارف - بنك - مصرف.
- حاسب - كمبيوتر - حاسوب - بي سي.
- هاتف نقال - محمول - جوال - موبايل - خليوي.
- مباراة - ماتش.
- فندق - أوتيل - هوتيل.

ومن الترادف الواجب مراعاته لدى محركات البحث الصور المختلفة للكيانات والأعلام:

الأمين العام للأمم المتحدة - سكرتير عام الأمم المتحدة - سكرتير الأمم المتحدة - بان كي مون.

الولايات المتحدة - أميركا - الولايات المتحدة الأميركية.

ومن الترادف الواجب مراعاته أيضا لدى محركات البحث الصور المختلفة للأعداد العربية وكتابتها بالحروف أو بالأرقام:

فلا شك في أن من يبحث عن الأعداد بالأرقام سيفقد نتائج كتبت بالحروف، وأن

من سيبحث عن الأعداد بالحروف سيفقد نتائج كتبت بالأرقام، ومن ثم وجب على محرك التعامل مع الشكلين في حال البحث عن أيهما.

- أحاديث الــ ٤٠ النووية - الأربعين النووية.
- ألف ليلة وليلة - ١٠٠٠ ليلة وليلة.
- العالم سنة ٢٠٠٣ - العالم سنة ألفين وثلاثة.

## ٢- مراعاة المشترك اللفظي (Homonymy):

وهو أن يكون للفظ الواحد أكثر من معنى دلالي، وينشأ عنه اللبس(Ambiguity)، ومثاله في العربية: العين، فهي في الأصل عضو الإبصار؛ والعين مكان يجري منه الماء؛ والعين: الوجيه في قوم، يقال: هو من أعيان الناس، والعين: الحسد؛ والعين مدينة في دولة الإمارات؛ وما إلى ذلك من معان أخرى.

ويتعين على محرك البحث عند اكتشاف اللبس من خلال آلياته أن يعرض الخيارات المختلفة، موجها سؤاله إلى المستخدم؛ أي المعاني تقصد ؟ ومن الممكن أيضا أن يعرض المحرك بعض الكلمات المتعلقة دلاليًّا بكلمة البحث، والتي من شأنها أن تزيل اللبس في كلمة البحث.

ومن ذلك البحث عن كلمة «العربية» -على سبيل المثال- فمن الممكن لمحرك البحث بعد اكتشاف اللبس في الكلمة أن يقترح بعض الحلول:

اللغة العربية.

قناة العربية.

المرأة العربية (صفة).

العربية (سيارة).

وكذا كلمة الجزيرة :

قناة الجزيرة.

شبه الجزيرة العربية.

جريدة الجزيرة.

ومن الواجب هنا تحديد المعنى المقصود بواسطة المستخدم، ويقع العبء الأكبر بعد ذلك على محرك البحث في اختيار النتائج وعرضها تبعًا للمعنى المراد من قبل المستخدم.

فالمعيار هنا إنما هو مدى تعامل المحرك مع قضية اللبس الدلالي من حيث اكتشافه وطرق تخلصه منه.

## ٣- مراعاة اللبس الصرفي على مستوى كلمات البحث وعلى مستوى كلمات التوقف:

**أولاً: على مستوى كلمات البحث:**

يظهر تأثيراللبس الصرفي جليًّا في محركات البحث على مستوى كلمات البحث[١] إذ تتعدد التوجيهات الصرفية للكلمة الواحدة، فيكون لكلمة البحث الواحدة تحليلات صرفية متعددة ومن ثم تحليلات دلالية متعددة، فمن شأن المحرك عند البحث عن كلمة «الدين» - على سبيل المثال - أن يؤدي عمله كحاطب ليل، وذلك بأن يجمع كل سياق لكلمة «الدين» بتحليلاتها الصرفية المختلفة دون نظر لمراد المستخدم أو مقصوده، فالبحث عن كلمة «الدين» يحتمل العديد من المعاني والدلالات التي يمكن أن يقصدها المستخدم، فالتحليلات الصرفية والدلالية لها متعددة وكثيرة ما بين المصدر والاسم الجامد وكونها جزءًا من العلم في التركيب «علاء الدين»، كما يلي:

| الكلمة المشكولة | السابق | الجذع | نوع الكلمة | الوزن | الجذر | قسم الكلم | اللاحق |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Voweled Word | Prefix | Stem | Type | Pattern | Root | POS Tags | Suffix |
| الدَّيْنُ | أل: التعريف | دين | اسم جامد | فَعْلُ | د ي ن | مفرد مذكر مرفوع في حالة التعريف | -- |
| الدِّينُ | أل: التعريف | دين | اسم جامد | فِعلُ | د ي ن | مفرد مذكر مرفوع في حالة التعريف | -- |

١ - راجع: اللبس في البنية والتركيب؛ أنماطه وأسبابه وآثاره وطرق التخلص منه، دراسة نحوية حاسوبية في العربية المعاصرة، عمرو جمعة عبد الرسول، رسالة دكتوراه بكلية دار العلوم، جامعة القاهرة، ٢٠١٤، ص ٧٤ وما بعدها.

| الكلمة المشكولة | السابق | الجذع | نوع الكلمة | الوزن | الجذر | قسم الكلم | اللاحق |
|---|---|---|---|---|---|---|---|
| Voweled Word | Prefix | Stem | Type | Pattern | Root | POS Tags | Suffix |
| الدَّيْنَ | أل: التعريف | دين | اسم جامد | فَعْلَ | د ي ن | مفرد مذكر منصوب في حالة التعريف | -- |
| الدِّينَ | أل: التعريف | دين | اسم جامد | فِعلَ | د ي ن | مفرد مذكر منصوب في حالة التعريف | -- |
| الدَّيْنِ | أل: التعريف | دين | اسم جامد | فَعْلِ | د ي ن | مفرد مذكر مجرور في حالة التعريف | -- |
| الدِّينِ | أل: التعريف | دين | اسم جامد | فِعلِ | د ي ن | مفرد مذكر مجرور في حالة التعريف | -- |
| الدِّين | -- | الدين | اسم علم (جزء من علم، مثل: علاء الدين) | -- | -- | -- | -- |
| الدِّينُ | أل: التعريف | دين | مصدر أصلي | فِعلُ | د ي ن | مفرد مذكر مرفوع في حالة التعريف | -- |
| الدَّيْنُ | أل: التعريف | دين | مصدر أصلي | فَعْلُ | د ي ن | مفرد مذكر مرفوع في حالة التعريف | -- |

| الكلمة المشكولة | السابق | الجذع | نوع الكلمة | الوزن | الجذر | قسم الكلم | اللاحق |
|---|---|---|---|---|---|---|---|
| Voweled Word | Prefix | Stem | Type | Pattern | Root | POS Tags | Suffix |
| الدِّينَ | أل: التعريف | دين | مصدر أصلي | فِعلَ | د ي ن | مفرد مذكر منصوب في حالة التعريف | -- |
| الدَّيْنَ | أل: التعريف | دين | مصدر أصلي | فَعْلَ | د ي ن | مفرد مذكر منصوب في حالة التعريف | -- |
| الدَّيْنِ | أل: التعريف | دين | مصدر أصلي | فَعْلِ | د ي ن | مفرد مذكر مجرور في حالة التعريف | -- |
| الدِّينِ | أل: التعريف | دين | مصدر أصلي | فِعلِ | د ي ن | مفرد مذكر مجرور في حالة التعريف | -- |

جدول رقم (١) لتحليل كلمة «الدين» بواسطة المحلل الصرفي «الخليل»

ومن ذلك أيضا استخدام محركات البحث في البحث عن كلمة «العلم» التي تحتمل من المعاني الصرفية والدلالية ما يلي[1]:

| الكلمة المشكولة | السابق | الجذع | نوع الكلمة | الوزن | الجذر | قسم الكلم | اللاحق |
|---|---|---|---|---|---|---|---|
| Voweled Word | Prefix | Stem | Type | Pattern | Root | POS Tags | Suffix |
| الْعِلْمُ | أل: التعريف | علم | اسم جامد | فِعْلُ | ع ل م | مفرد مذكر مرفوع في حالة التعريف | -- |
| الْعَلَمُ | أل: التعريف | علم | اسم جامد | فَعَلُ | ع ل م | مفرد مذكر مرفوع في حالة التعريف | -- |
| الْعِلْمَ | أل: التعريف | علم | اسم جامد | فِعْلَ | ع ل م | مفرد مذكر منصوب في حالة التعريف | -- |
| الْعَلَمَ | أل: التعريف | علم | اسم جامد | فَعَلَ | ع ل م | مفرد مذكر منصوب في حالة التعريف | -- |
| الْعِلْمِ | أل: التعريف | علم | اسم جامد | فِعْلِ | ع ل م | مفرد مذكر مجرور في حالة التعريف | -- |
| الْعَلَمِ | أل: التعريف | علم | اسم جامد | فَعَلِ | ع ل م | مفرد مذكر مجرور في حالة التعريف | -- |

١ - راجع: اللبس في البنية والتركيب؛ أنماطه وأسبابه وآثاره وطرق التخلص منه، دراسة نحوية حاسوبية في العربية المعاصرة، عمرو جمعة عبد الرسول، ص ٧٧ وما بعدها.

| الكلمة المشكولة | السابق | الجذع | نوع الكلمة | الوزن | الجذر | قسم الكلم | اللاحق |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Voweled Word | Prefix | Stem | Type | Pattern | Root | POS Tags | Suffix |
| الْعِلْمُ | أل: التعريف | علم | مصدر أصلي | فِعْلُ | ع ل م | مفرد مذكر مرفوع في حالة التعريف | -- |
| الْعَلْمُ | أل: التعريف | علم | مصدر أصلي | فَعْلُ | ع ل م | مفرد مذكر مرفوع في حالة التعريف | -- |
| الْعَلَمُ | أل: التعريف | علم | مصدر أصلي | فَعَلُ | ع ل م | مفرد مذكر مرفوع في حالة التعريف | -- |
| الْعَلَمَ | أل: التعريف | علم | مصدر أصلي | فَعَلَ | ع ل م | مفرد مذكر منصوب في حالة التعريف | -- |
| الْعَلْمَ | أل: التعريف | علم | مصدر أصلي | فَعْلَ | ع ل م | مفرد مذكر منصوب في حالة التعريف | -- |
| الْعِلْمَ | أل: التعريف | علم | مصدر أصلي | فِعْلَ | ع ل م | مفرد مذكر منصوب في حالة التعريف | -- |
| الْعَلْمِ | أل: التعريف | علم | مصدر أصلي | فَعْلِ | ع ل م | مفرد مذكر مجرور في حالة التعريف | -- |
| الْعَلَمِ | أل: التعريف | علم | مصدر أصلي | فَعَلِ | ع ل م | مفرد مذكر مجرور في حالة التعريف | -- |

| الكلمة المشكولة | السابق | الجذع | نوع الكلمة | الوزن | الجذر | قسم الكلم | اللاحق |
|---|---|---|---|---|---|---|---|
| Voweled Word | Prefix | Stem | Type | Pattern | Root | POS Tags | Suffix |
| الْعِلْمِ | أل: التعريف | علم | مصدر أصلي | فِعْلِ | ع ل م | مفرد مذكر مجرور في حالة التعريف | -- |

جدول رقم (٢) لتحليل كلمة «العلم» بواسطة المحلل الصرفي «الخليل»

ومنه أيضا البحث عن كلمة «التهم» التي تحتمل الاسم والفعل في الوقت نفسه وفقا لتحليلات الصرفية على محلل الخليل الصرفي[١]:

| الكلمة المشكولة | السابق | الجذع | نوع الكلمة | الوزن | الجذر | قسم الكلم | اللاحق |
|---|---|---|---|---|---|---|---|
| Voweled Word | Prefix | Stem | Type | Pattern | Root | POS Tags | Suffix |
| التُّهَمُ | أل: التعريف | تهم | مصدر أصلي | فُعَلُ | ت ه م | جمع تكسير مرفوع في حالة التعريف | -- |
| التُّهَمَ | أل: التعريف | تهم | مصدر أصلي | فُعَلَ | ت ه م | جمع تكسير منصوب في حالة التعريف | -- |

١ - راجع: اللبس في البنية والتركيب؛ أنماطه وأسبابه وآثاره وطرق التخلص منه، دراسة نحوية حاسوبية في العربية المعاصرة، عمرو جمعة عبد الرسول، ص ٨١ وما بعدها.

| الكلمة المشكولة | السابق | الجذع | نوع الكلمة | الوزن | الجذر | قسم الكلم | اللاحق |
|---|---|---|---|---|---|---|---|
| Voweled Word | Prefix | Stem | Type | Pattern | Root | POS Tags | Suffix |
| التُّهَمِ | أل: التعريف | تهم | مصدر أصلي | فُعَلِ | ت ه م | جمع تكسير مجرور في حالة التعريف | -- |
| الْتَهَمَ | -- | التهم | فعل ماض مبني للمعلوم | افْتَعَلَ | ل ه م | ثلاثي مزيد مسند إلى الغائب (هو) متعد | -- |
| الْتُهِمَ | -- | التهم | فعل ماض مبني للمجهول | افْتُعِلَ | ل ه م | ثلاثي مزيد مسند إلى الغائب (هو) متعد | -- |
| الْتَهِمْ | -- | التهم | فعل أمر | افْتَعِلْ | ل ه م | ثلاثي مزيد مسند إلى المخاطَب أنت متعد | -- |

جدول رقم (٣) لتحليل كلمة «التهم» بواسطة المحلل الصرفي «الخليل»

**ثانيا: على مستوى كلمات التوقف:**

فإن من المتفق عليه في نظم البحث والاسترجاع أن يتم استبعاد بعض كلمات البحث عند القيام بعملية البحث[1] وذلك لعدم دلالتها على معان في حد ذاتها، وللإسهام في الإسراع بعملية البحث، وهي ما يطلق عليها كلمات التوقف (Stop

١ – ما لم توضع مع كلمات أخرى بين علامتي اقتباس " "، ومثاله: "ليلة في حب مصر".

Words) أو كلمات الاستبعاد، وهي كلمات تتكرر كثيرًا في النصوص، ومن أمثلتها: الضمائر، وحروف العطف، وحروف الجر. وتسعى معظم نظم استرجاع المعلومات إلى عمل قائمة بالكلمات التي يجب تجاهلها عند إجراء عملية البحث، وتستخدم بعض النظم أساليب إحصائية لتحديد أكثر هذه الكلمات شيوعًا من حيث استخدامها. ففي الإنجليزية – على سبيل المثال – تسع عشرة كلمة تتوزع بين أدوات التعريف وحروف الجر والعطف، وهي: ( ,a, am, and an, as, at, but, by, for, in, into, no, not, of, on, or, the, to us) [1]

أما كلمات التوقف في العربية، فيمكن أن تشمل حروف المعاني، وأسماء الإشارة، والأسماء الموصولة، وأدوات الشرط، وأدوات الاستفهام، ... إلى آخر هذه الكلمات[2].

ويتجلى تأثير اللبس الصرفي في نظم استرجاع المعلومات عند استخدام كلمات ملبسة ضمن كلمات التوقف «Stop Words»، فكلمات مثل: «بين»، «عن»، «من»، «على»[3]،

١ – راجع: قواعد البيانات المحملة على الأقراص المدمجة: دراسة حالة لقاعدة بيانات الإنتاج الفكري الإسلامي INDEX ISLAMICUS ON CD-ROM، هاشم فرحات، مجلة مكتبة الملك فهد الوطنية، مجلد ٦، عدد ٢ (رجب ـ ذو الحجة ١٤٢١هـ/ أكتوبر ٢٠٠٠ـ مارس ٢٠٠١م)، ص ٢٠ على الرابط
http://faculty.ksu.edu.sa/Alhendawy/My%20publications/%D9%82%D9%88%D8%A7%D8%B9%D8%AF%20%D8%A7%D9%84%D8%A8%D9%8A%D8%A7%D9%86%D8%A7%D8%AA.pdf
في: 12-12-2013.

٢ – تتوافر على الشبكة العالمية للمعلومات (الإنترنت) بعض المصادر المفتوحة التي أعدت قائمة بكلمات التوقف العربية ومنها موقع عرب آيز (وهو موقع تشاركي مفتوح للجميع دون قيود، يوفر مناخا للنقاش وتبادل الخبرات والمعرفة حول اللغة العربية ولسانياتها وما يتعلّق بدعمها والارتقاء بها في الحوسبة بشكل عام، والبرمجيات الحرة على وجه الخصوص).
وتحتوي القائمة – التي أطلق عليها «مستبعدات الفهرسة» – ٦٧٢ كلمة توقف تشمل: الاتجاهات، وأدوات الاستثناء، وأسماء الإشارة، وأسماء الاستفهام، وأسماء الشرط، والأسماء الموصولة، والحروف الأبجدية، وأسماء الشهور، وأسماء الأصوات، وأسماء الأعداد، وأسماء الأفعال، والأسماء الخمسة، وأفعال التعجب، والأفعال الخمسة، وأفعال القلوب، وإن وأخواتها، وأدوات التوكيد، والكلمات الجامدة، وحروف العطف المنفصلة، وبعض الحروف والإشارات، وبعض الحروف والأفعال والأدوات، والضمائر المتصلة المجرورة والمنصوبة، والضمائر المنفصلة، والظروف، وأسماء العملات، وكاد وأخواتها، وكان وأخواتها، والأفعال المتعدية لمفعولين، والأفعال المتعدية لثلاثة مفعولات، وأدوات المدح والذم، والمشبهات بليس، وبعض المنصوبات، ولا النافية للجنس، وأسماء الأيام. مع إضافة القواعد اللغوية لتوليد هذه صور الكلمات مع سوابقها (العطف والجر والاستفهام والتعريف) ولواحقها (الإضافة أو الإسناد للضمائر والتصريف والاشتقاق) والتي بلغ عددها (١٢٢٠١) صورة.
راجع: الرابط في ١٢/ ١٢/ ٢٠٠٩
http://wiki.arabeyes.org/مستبعدات_الفهرسة

٣ – حرف الجر «على» قد يلتبس مع العلم «عَلِيّ» في الكتابة المصرية فقط، ومن هنا فاللبس فيها يقتصر على الكتابة المصرية،

«قبل»، «بعد»، «عبر»، «غير»، «مهما»، «ظل»، «هم»، «حول»، «دون»، «قد»[1] لها تحليل صرفي آخر مغاير لكونها من كلمات التوقف، ومن ثم فإن تضمينها كلمات التوقف أمر غير دقيق لغويا، ويؤدي لتجاهل العديد من النتائج المهمة في البحث.

فكلمة مثل «قد» – على سبيل المثال –يمكن أن يكون من تحليلاتها الصرفية ما يلي[2]:

| الكلمة المشكولة | السابق | الجذع | نوع الكلمة | الوزن | الجذر | قسم الكلم | اللاحق |
|---|---|---|---|---|---|---|---|
| Voweled Word | Prefix | Stem | Type | Pattern | Root | POS Tags | Suffix |
| قَدُّ | -- | قد | اسم جامد | فَعُّ | ق د د | مفرد مذكر مرفوع في حالة الإضافة | -- |
| قَدٌّ | -- | قد | اسم جامد | فَعٌّ | ق د د | مفرد مذكر مرفوع نكرة | -- |
| قَدَّ | -- | قد | اسم جامد | فَعَّ | ق د د | مفرد مذكر منصوب في حالة الإضافة | -- |
| قَدِّ | -- | قد | اسم جامد | فَعِّ | ق د د | مفرد مذكر مجرور في حالة الإضافة | -- |

أما الكتابة الشامية فتستخدم الياء المنقوطة (ي) للتفريق بين الياءين (عَلِيّ).

١ – راجع: اللغة العربية المعاصرة في المجال السياسي، دراسة لغوية حاسوبية من واقع ذخيرة لغوية، هبة سيد راشد، رسالة ماجستير، كلية الألسن، جامعة عين شمس، ٢٠٠٨، ص ٥٨.

٢ – راجع: اللبس في البنية والتركيب؛ أنماطه وأسبابه وآثاره وطرق التخلص منه، دراسة نحوية حاسوبية في العربية المعاصرة، عمرو جمعة عبد الرسول، ص ٧٠ وما بعدها.

| الكلمة المشكولة<br>Voweled Word | السابق<br>Prefix | الجذع<br>Stem | نوع الكلمة<br>Type | الوزن<br>Pattern | الجذر<br>Root | قسم الكلم<br>POS Tags | اللاحق<br>Suffix |
|---|---|---|---|---|---|---|---|
| قَدٍّ | -- | قد | اسم جامد | فَعٍّ | ق د د | مفرد مذكر مجرور نكرة | -- |
| قَدُّ | -- | قد | مصدر أصلي | فَعُّ | ق د د | مفرد مذكر مرفوع في حالة الإضافة | -- |
| قَدٌّ | -- | قد | مصدر أصلي | فَعٌّ | ق د د | مفرد مذكر مرفوع نكرة | -- |
| قَدَّ | -- | قد | مصدر أصلي | فَعَّ | ق د د | مفرد مذكر منصوب في حالة الإضافة | -- |
| قَدِّ | -- | قد | مصدر أصلي | فَعِّ | ق د د | مفرد مذكر مجرور في حالة الإضافة | -- |
| قَدٍّ | -- | قد | مصدر أصلي | فَعٍّ | ق د د | مفرد مذكر مجرور نكرة | -- |
| قَدَّ | -- | قد | فعل ماض مبني للمعلوم | فَعَّ | ق د د | ثلاثي مجرد مسند إلى الغائب(هو) متعد | -- |

| الكلمة المشكولة Voweled Word | السابق Prefix | الجذع Stem | نوع الكلمة Type | الوزن Pattern | الجذر Root | قسم الكلم POS Tags | اللاحق Suffix |
|---|---|---|---|---|---|---|---|
| قُدَّ | -- | قد | فعل ماض مبني للمجهول | فُعَّ | ق د د | ثلاثي مجرد مسند إلى الغائب(هو) متعد | -- |
| قِدْ | -- | قد | فعل أمر | فِلْ | ق و د | ثلاثي مجرد مسند إلى المخاطَب أنت متعد | -- |
| قُدَّ | -- | قد | فعل أمر | فُعَّ | ق د د | ثلاثي مجرد مسند إلى المخاطَب أنت متعد | -- |
| قُدْ | -- | قد | فعل أمر | فُلْ | ق و د | ثلاثي مجرد مسند إلى المخاطَب أنت متعد | -- |
| قِدْ | -- | قد | فعل أمر | عِلْ | و ق د | ثلاثي مجرد مسند إلى المخاطَب أنت متعد | -- |

| اللاحق | قسم الكلم | الجذر | الوزن | نوع الكلمة | الجذع | السابق | الكلمة المشكولة |
|---|---|---|---|---|---|---|---|
| Suffix | POS Tags | Root | Pattern | Type | Stem | Prefix | Voweled Word |
| -- | -- | -- | -- | حَرف تَحقِيق وتَقرِيب | قد | -- | قَدْ |

جدول رقم (٤) لتحليل كلمة «قد» بواسطة المحلل الصرفي «الخليل»

وهي الاحتمالات الصرفية التي يصعب معها عدّها مع ما يشبهها من كلمات التوقف التي يجب على محلل البحث ذي السمات اللغوية العربية إهمالها عند فهرسة النصوص العربية والبحث فيها.

والخلاصة أن البحث عن إحدى هذه الكلمات الملبسة صرفيا باستخدام محركات البحث يقلل من حجم النتائج الصحيحة لعملية البحث حيث تحتمل الكلمة حينئذ العديد من المعاني التي تتسبب في إخراج المحرك للعديد من النتائج غير الدقيقة؛ لاختلاف التحليلات الصرفية لكلمات البحث هذه. ولذا يمكن اعتبار مدى مراعاة محرك البحث للبس الصرفي والدلالي معيارًا مهمًّا من معايير تقييم هذا المحرك والحكم عليه.



## المبحث الثالث

## المعايير اللغوية والفنية لتقييم برامج التعرف الضوئي على الحروف العربية

### تمهيد:

تعد عملية رقمنة مصادر المعلومات العربية أمرًا بالغ الأهمية؛ إذ تُحوَّل من خلالها نصوص مصادر المعلومات الورقية إلى شكل إلكتروني قابل لعمليات المعالجة الآلية، وذلك من خلال المسح الضوئي للنصوص، ومن ثم الوصول إلى نسخة إلكترونية مطابقة تماماً للنص الأصلي الموجود في صورة ملفات jpg . أو pdf. على سبيل المثال.

ومع ازدياد الحاجة إلى رقمنة مثل هذه الملفات، من خلال التقنيات المتخصصة في التعرف الضوئي على الحروف Optical Character Recognition «OCR" زادت الحاجة إلى تحديد معايير لغوية لتقييم هذه البرامج.

والمبحث الذي بين أيدينا يرسم خطوطا عريضة للمعايير اللغوية - على وجه الخصوص - لتقييم مثل هذه البرامج، وتتباين هذه المعايير ما بين الفنية واللغوية؛ فالفنية منها تخص شكل الحرف العربي وأنواع خطوطه المختلفة سواء باليد أو بالآلة الكاتبة، أما المعايير اللغوية فتخص الجانب اللغوي للحروف العربية من حيث اتصالها وانفصالها والبدء بها أو الانتهاء بها، إضافة إلى ضبط الحروف العربية بالشكل في المستوى الكلاسيكي منها، وهذا ما يزيد من تعقيد عملية التعرف على الحروف العربية،

ومن المعايير اللغوية كذلك مدى استخدام القواعد اللغوية في التعرف على الحروف العربية، أو استخدام آلية لغوية لتصحيح الأخطاء اللغوية في التعرف. وقبل أن نشرع في تفصيل هذه المعايير، نعرض أولا لمشكلات التعرف على الحروف العربية.

### مشكلات التعرف على الحروف العربية، وتكمن في:

الكتابة المتصلة للحروف العربية: حيث تتميز العربية بكتابة حروفها متصلة غير منفصلة مثل الإنجليزية مثلاً، ومن ثم يصعب التعرف عليها نتيجة لتداخلها.

نقط الحروف أو الإعجام: تشترك نصف حروف العربية في نقط الحروف أو إعجامها، فالنقط مهم جدا في الأبجدية العربية، وذلك أنه يفرّق بين الحروف، والعرب قديمًا كانوا يطلقون عليه الإعجام أي إزالة الغموض ويأتي في مقابله الإهمال، فالحروف إما مهملة أو معجمة. ولكن هذا النقط أو الإعجام يمثل صعوبة كبيرة عند عملية التعرف الآلي على الحروف العربية لتشابه الكثير من الحروف العربية في النقط. فالحروف العربية (ج،ح،خ)، لا يكاد يفرق بينها إلا التنقيط، فحروف كالجيم والحاء والخاء تكتب هكذا (ح) مع اختلاف وضع النقطة على الحرف أو عدم وضعها، وهو ما يضاف إلى صعوبات التعرف الآلي على الحروف.

**والجدول الآتي يرصد بعض صور هذه الحروف واحتمالاتها عند الرسم الكتابي:**

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| احتمال ١ | | ب | ت | ـجـ | د | ر | ـشـ | ـصـ | ـطـ | ـعـ | ـفـ | ـه | ي |
| احتمال ٢ | | يـ | ن | ـحـ | ذ | ز | ـثـ | ـضـ | ـظـ | ـغـ | ـقـ | ـة | ى |
| احتمال ٣ | | | | ـخـ | | | ـتـ | | | ـفـ | ـعـ | | |
| احتمال ٤ | | | | | | | | | | ـقـ | ـغـ | | |

جدول رقم (٥) يوضح بعض صور الحروف المعجمة والمهملة في اللغة العربية.

الضبط بالشكل: تتميز العربية بالتعبير عن الأصوات بحركات التشكيل، وتأتي هذه الحركات مقابلة للصوائت في اللغات الأخرى (e-o-i)، فحركات الكسرة والضمة

والفتحة والسكون ( ُ َ ْ ) كلها علامات على الصوائت وأدلة على وجودها. وهكذا يعتور الكلمة العربية الكثير من الزوائد الشكلية ابتداءً من نقط الحروف وانتهاءً بحركات الضبط بالشكل.

الخطوط الفنية الزخرفية (كوفي – أندلسي – فارسي – ...): تتميز العربية بإمكان كتابة حروفها بعدد من الحروف الزخرفية، التي تضيف إلى الصعوبات السابقة المتعلقة بخصائص الحروف العربية صعوبة أخرى.

تغير أشكال الحروف العربية بتغير مواقعها وتغير أشكالها بتغير نوع الخط المكتوبة به فحرف (س) على سبيل المثال يكتب في أول الكلام (سـ) وفي آخر الكلمة (س) وفي وسط الكلمة (ـسـ)، وكذا يختلف شكله بخط النسخ عنه في حال خط الرقعة.

ينبغي كذلك أن نوضح أن ثمة مستويين للنصوص العربية المعالجة، أولها مستوى النصوص غير المضبوطة بالشكل، وهي نصوص بسيطة. وثانيها مستوى النصوص المضبوطة بالشكل، وهي نصوص معقدة.

**أما المعايير اللغوية والفنية لتقييم برامج التعرف الآلي على الحروف العربية فهي على النحو الآتي:**

### المعيار الأول:

التعرف على جميع حروف الخطوط العربية المستخدمة فى الكتابة العربية، ولذا يقترح الباحث مجموعة من المصادر، لتكون مصادر خطوط نموذجية للنصوص العربية (من حروف، وأرقام، وعلامات تشكيل، وعلامات ترقيم) لاستخدامها في عملية تقييم عملية التعرف الآلي على الحروق العربية؛ إذ يجب أن تضم:

(١) مجموعة من خطوط اليد لمجموعة متفاوتة في السن وفي التعليم.

(٢) مجموعة مختلفة الخطوط من الصحف والجرائد في شتى أنحاء الوطن العربي.

(٣) مستندات مكتوبة على الآلة الكاتبة.

(٤) نصوص لخطاطين بشتى أنواع الخط العربي غير الزخرفية.

(٥) نماذج من خطوط (الأوفيس) غير الزخرفية.

(٦) تنسيقات مختلفة لصور الخطوط السابقة جميعها. فاحتواء النص على تنسيقات

غاية في التعقيد -مثل: (وجود عدة أعمدة رأسية، ووجود إيضاحات أو هوامش أو حواشٍ في أماكن غير منتظمة، وغير ذلك)- سيكون له تأثير سلبي على جودة التعرف الضوئي على الحروف.

(٧) نصوص متعددة اللغات أو متعددة الرموز أو الحروف (حروف وأرقام متداخلة، أو تداخل العديد من اللغات في النصوص التي تحوي أكثر من لغة).

**وسنحاول في الصفحات القليلة القادمة التمثيل لكل من النماذج السابقة:**

## (١) مجموعة من خطوط اليد لمجموعة متفاوتة في السن وفي التعليم.

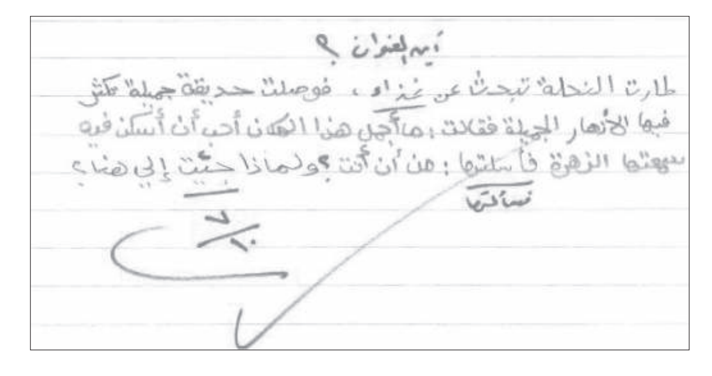
دمشقُ .. مهرجانُ الماء والياسمين

نزار قباني

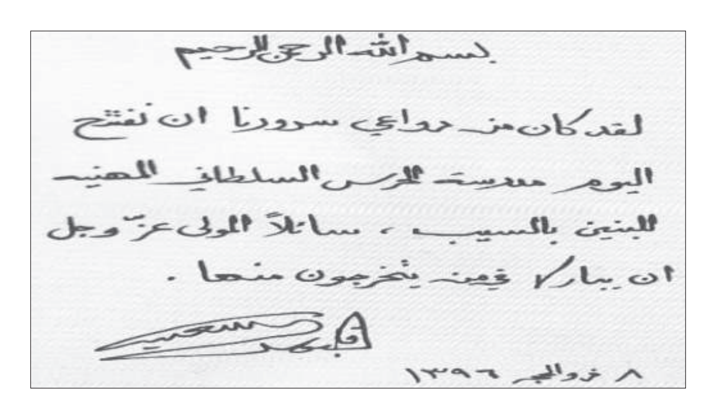
١

لا أستطيعُ أن أكتبَ عن دمشق، دونَ أن يُعرّشَ الياسمينُ
على أصابعي.
ولا أستطيعُ أن أنطقَ اسمَها، دونَ أن يكتظَّ فمي بعصير
المشمشِ، والرمّانِ، والتوت، والسفرجل.
ولا أستطيعُ أن أتذكّرها، دونَ أن تحطَّ على جدار ذاكرتي
ألفُ حمامةٍ.. وتطيرَ ألفُ حمامة..

شكل رقم (١) يمثل نموذجا لخط اليد


أين العنوان ؟
طارت النحلة تبحث عن غذاء، فوصلت حديقة جميلة تكثر
فيها الأزهار الجميلة فقالت: ما أجمل هذا المكان أحب أن أسكن فيه
سمعتها الزهرة فسألتها: من أنت؟ ولماذا جئت إلى هنا؟
فسألتها
٧/١٠

شكل رقم (٢) يمثل نموذجا لخط اليد



بسم الله الرحمن الرحيم

لقد كان من دواعي سرورنا ان نفتتح اليوم مدرسة الحرس السلطاني العماني للبنين بالسيب ، سائلاً المولى عز وجل ان يبارك فيمن يتخرجون منها .

٨ ذو الحجة ١٣٩٦

شكل رقم (٣) يمثل نموذجا لخط اليد

## (٢) مجموعة مختلفة الخطوط من الصحف والجرائد في شتى أنحاء الوطن العربي.

وقالت المصادر إن الجيش الباكستانى اعتقل «أبودجانة» أثناء الحملة التى قامت بها قوات الجيش على منطقة القبائل فى وادى سوات الشهور الماضية، مشيرة إلى مشاركة بعض أجهزة الاستخبارات الأجنبية والعربية فى تلك الحملة.

شكل رقم (٤) يمثل نموذجا لخط إحدى الصحف العربية

**أكبر علم فلسطيني في العالم**

**في طريقه إلى موسوعة «غينيس» للأرقام القياسية**

الخليل-وفا- يعتزم مجموعة من الشباب الفلسطينيين من الضفة الغربية وقطاع غزة والشتات، تنفيذ مشروع أكبر علم فلسطيني في العالم بطول ٢٠٠ متر وعرض ٦٣ متراً وبزنة طنين، حيث سيسجل هذا العلم ضمن الأرقام القياسية في موسوعة " غينيس " العالمية.

ولفت جلال مخارزة، رئيس غرفة تجارة جنوب الخليل إلى أن فكرة تنفيذ هذا المشروع جاءت عبر مجموعة من الشباب في الداخل والخارج ( نحو ٤٠ شاباً وشابة ) يتواصلون ببعضهم عبر شبكة المعلومات العنكبوتية "الإنترنت".

وأوضح أنه عندما طرحت هذه الفكرة الأعمال الساعدة في دعم هذه الفكرة التي تعتبر سابقة، فيما يتعلق بالأرقام المسجلة في موسوعة " غينيس " حول أحجام أعلام دول العالم المختلفة، إذ أن هذا العلم سوف يكون بمثابة إنجاز فلسطيني يلفت الانتباه إلى القضية والمشروع التحرري الفلسطيني وإثبات للذات.

واعتبر مخارزة، "أن ذلك هو ضمن المساعدات التي تقدمها غرفة تجارة جنوب الخليل لتفعيل دور الشباب في المجتمع، ومساعدتهم في الهروب من أجواء البطالة وقلة فرص العمل، وإيجاد وسائل ومشاريع يمكن من خلالها أن يعبروا عما يمتلكونه من قدرات مختلفة، إضافة إلى أنه نوع من شد الانتباه الفلسطينيين في لبنان، وذلك للفت الانتباه لمعاناة اللاجئين في مخيمات الشتات وما يعانونه من ظروف معيشية صعبة للغاية، والتفكير في الأسباب التي أدت الى طردهم من وطنهم واللجوء إلى الخارج.

وأضاف أنه سيتم عرض العلم كذلك في بعض الدول العربية والأجنبية، علماً بأن هذا العلم يحتاج لنحو ٦٠٠ شخص لحمله والسير فيه، لافتاً إلى أنه تم الاطلاع على الأرقام القياسية المسجلة لدى موسوعة " غينيس " بهذا الخصوص، حيث تبين أن الرقم القياسي المسجل لأكبر علم في العالم هو علم البرازيل، ولكن بحجم أقل من العلم الفلسطيني المزمع إنجازه.

شكل رقم (٥) يمثل نموذجا لخط إحدى الصحف العربية

توجه لاختيار ابو مازن رئيسا للحركة بالتزكية.. تمديد الاجتماعات الى الثلاثاء

**مؤتمر فتح يشهد نقاشا «أكثر من عاصف» حول سقوط غزة:**

**دحلان يتبادل الاتهامات مع قريع ويحمل القيادة المسؤولية كاملة**

شكل رقم (٦) يمثل نموذجا لخط إحدى الصحف العربية

الدوحة - إيمان الشمري

ضمن فعاليات يوم الكتاب العالمي المقامة في مدرسة قطر الإعدادية نظم أخيرا معرض للكتاب في القاعة الرياضية للمدرسة. المعرض أقيم لتعزيز ثقافة الفتيات وتشجيعهن على حب القراءة والاطلاع لفوائده الكبيرة في بناء وصقل الشخصية المثقفة وتماشيا مع رسالة المدرسة التي من ضمن أهدافها الرئيسية بناء الشخصية المثقفة الواعية المتوازنة. يوم الكتاب العالمي كانت له أصداؤه الواسعة، وكان حقا أن تقام له فعالية على أرض الإبداع، أرض قطر الإعدادية.

**المعرض حديقة غناء بزهور الثقافة**

كانت دور النشر المشاركة بكتبها كالآتي- دار قطري بن الفجاءة، دار الثقافة للنشر والتوزيع ومكتبة النور.

أما تنوع الكتب فكان محيرا لبعض الزوار الذين اعتبروا التجول بين أركان المعرض بمثابة المرور بحديقة غناء يحتار القارئ ماذا ينتقي وماذا يترك! كانت أغلب الكتب التي لاقت قبولا هي الكتب الدينية للمشايخ المعروفين أمثال الدكتور محمد العريفي وعائض القرني وكذلك الكتب التطويرية والتعليمية للطالبات، وكتب تعليم الطبخ والروايات الإنجليزية المترجمة للعربية أمثال قصص الكاتبة «أجاثا كريستي» البوليسية، والروايات العربية للكتاب العرب أمثال نجيب محفوظ قد لاقت إقبالا كبيرا من الزوار من المدارس الثانوية.

**الفعاليات المصاحبة**

أقيمت فعاليات مصاحبة للمعرض مثل ورشة «كيف أصنع بيدي» من إعداد الطالبات، تهدف إلى تعليم الطالبات كيفية صنع بعض المشغولات اليدوية المفيدة في حياتهن اليومية مثل تزيين الدفاتر والملفات والأدوات المكتبية. كما كان هناك ركن للمأكولات الشعبية والمشروبات السريعة للزوار

**للشعر نكهة مميزة شاعرية**

عبق المشاعر وصدقها الذي ترجم إلى كلمات شاعرية أصبحت قصائد أضافت نكهة رائعة لأجواء المعرض. حيث أحييت الطالبتان الموهوبتان شعريا جفلة المضاحكة وهيفاء الكواري من الصف الثالث الإعدادي أصبوحة شعرية رائعة- بدأت بالتعريف عن أنفسهن ومن ثم إلقاء قصائد شعرية. قرأت الموهوبة جفلة أولا قصيدة عن قطر ذكرت فيها تاريخ قطر المجيد وإنجازات الشيخ حمد آل ثاني ونهضتها المميزة بين الدول الباقية.

أما الطالبة هيفاء الكواري فقد انفردت بقصيدة مميزة عن الأصحاب والأصدقاء تحدثت بها عن منظورها

شكل رقم (٧) يمثل نموذجا لخط إحدى الصحف العربية

## (٣) مستندات مكتوبة على الآلة الكاتبة.

إذا كان الخط حسن الوصف، مليح الرصف، مفتوح العيون، أملس المتون، كثير الإئتلاف، قليل الإختلاف، هشت إليه النفوس، واشتهته الأرواح، حتى إن الإنسان ليقرأه ولو كان فيه كلام رديء، ومعنى رديء مستزيد فيه ولو كثر، من غير سآمة تلحقه، وإن كان الخط قبيحا، مجته الأفهام ولفظته العيون والنفوس، وسئم قارئه، وإن كان فيه من الحكمة عجائبها ومن الألفاظ غرائبها. (٢٦)

شكل رقم (٨) يمثل نموذجا لخط الآلة الكاتبة

## (٤) نصوص لخطاطين بشتى أنواع الخط العربي غير الزخرفية:

### الخط الرقعي

عيناك غابتا نخيل ساعة السحر ...
أو شرفتان راح ينأى عنهما القمر
عيناك حين تبسمان تورق الكروم
وترقص الأضواء كالأقمار في نهر
يرجه المجداف وهنا ساعة السحر
كأنما تنبض في غوريهما، النجوم ...

شكل رقم (٩) يمثل نموذجا لخط الرقعة مضبوطًا بالشكل

## (٥) نماذج من خطوط الأوفيس غير الزخرفية.

نماذج من خطوط الأوفيس غير الزخرفية.

| Office fonts | | | |
|---|---|---|---|
| DecoType Naskh | بسم الله الرحمن الرحيم | Microsoft Uighur | بسم الله الرحمن الرحيم |
| DecoType Naskh Extensions | بسم الله الرحمن الرحيم | Monotype Koufi | بسم الله الرحمن الرحيم |
| DecoType Naskh Special | بسم الله الرحمن الرحيم | Mudir MT | بسم الله الرحمن الرحيم |
| DecoType Naskh Swashes | بسم الله الرحمن الرحيم | Simplified Arabic | بسم الله الرحمن الرحيم |
| DecoType Naskh Variants | بسم الله الرحمن الرحيم | Simplified Arabic Fixed | بسم الله الرحمن الرحيم |
| DecoType Thuluth | بسم الله الرحمن الرحيم | Tahoma | بسم الله الرحمن الرحيم |
| Farsi Simple Bold | بسم الله الرحمن الرحيم | Times New Roman | بسم الله الرحمن الرحيم |
| Farsi Simple Outline | بسم الله الرحمن الرحيم | Traditional Arabic | بسم الله الرحمن الرحيم |
| Arabic Typesetting | بسم الله الرحمن الرحيم | Akhbar MT | بسم الله الرحمن الرحيم |
| Arial | بسم الله الرحمن الرحيم | Andalus | بسم الله الرحمن الرحيم |

جدول رقم (٦) لخطوط الأوفيس الزخرفية منها وغير الزخرفية

**(٦) تنسيقات مختلفة لصور الخطوط السابقة جميعها، فاحتواء النص على تنسيقات غاية في التعقيد - مثل (وجود عدة أعمدة رأسية، ووجود إيضاحات أو هوامش أو حواشٍ في أماكن غير منتظمة، وغير ذلك)- سيكون له تأثير سلبي على جودة التعرف الضوئي على الحروف.**

جا حـ + جا ى = ٢ جا ١/٢ (حـ + ى) جتا ١/٢ (حـ − ى)
جتا حـ + جتا ى = ٢ جتا ١/٢ (حـ + ى) جتا ١/٢ (حـ − ى)
جا حـ − جا ى = ٢ جتا ١/٢ (حـ + ى) جا ١/٢ (حـ − ى)
جتا حـ − جتا ى = − ٢ جا ١/٢ (حـ + ى) جا ١/٢ (حـ − ى)
جا ٢ حـ = ٢ جا حـ جتا حـ
ظا ٢ حـ = ٢ ظا حـ / (١ − ظا$^{٢}$ حـ)
جتا ٢ حـ = جتا$^{٢}$ حـ − جا$^{٢}$ حـ
= ٢ جتا$^{٢}$ حـ − ١
= ١ − ٢ جا$^{٢}$ حـ

شكل رقم (١٠) يمثل نموذجا لاحتواء النص على تنسيقات غاية في التعقيد

## (٧) نصوص متعددة اللغات أو متعددة الرموز أو الحروف

**(حروف وأرقام متداخلة ، أو تداخل العديد من اللغات في النصوص التي تحوي أكثر من لغة).**

تَعْيِينُ الْهُوِيَّةِ أو الشَّخْصِيَّة identification
تَغَابَى عن to ignore, disregard, overlook, be unmindful of
تَغَابَنَ to wrong one another; to cheat one another
تَغَازَلَ to court or woo one another; to flirt with one another
تَغَاضٍ (التَّغَاضِي) (عن) overlooking, disregard(ing), condonation, connivance, winking at, blinking at, shutting one's eyes to; excuse, forgiveness, pardon(ing)
تَغَاضَى عن to overlook, disregard, condone, connive at, wink at, blink at, bypass, pass over, let pass, shut one's eyes to; to excuse, forgive, pardon
تَغَاضَى عَيْنَهُ أو عَيْنَيْه to close one's eyes
تَغَافَلَ عن – راجع أَغْفَلَ
تَغَافَلَ: تَظَاهَرَ بِالغَفْلَة to feign inattention or inadvertence, feign negligence or carelessness, pretend to be inattentive or inadvertent, pretend to be negligent or careless

alimented; to feed on, live on; to eat; to be fed (with), charged (with), provided (with), supplied (with)
تَغْذَوِيّ nutritional
تَغْذِيَة (شَخْصٍ) nutrition, feeding, nourishment, nourishing, alimentation
تَغْذِيَة (آلَةٍ، فُرْنٍ إلخ) feed(ing), charging; providing with, supplying with
تَغْذِيَةٌ اسْتِرْجَاعِيَّةٌ أو راجِعَة feedback
عِلْمُ التَّغْذِيَة dietetics, sitology
تَغَرَّى to be glued, fixed with glue, conglutinated, agglutinated; to conglutinate, agglutinate
تَغَرَّبَ: نَزَحَ، هاجَرَ – راجع اِغْتَرَبَ
تَغَرُّب – راجع اِغْتِراب
تَغَرَّدَ: غَرَّدَ to sing, warble
تَغَرَّضَ لِـ: تَحَيَّزَ to take sides with, side with, be partial to, be biased for, have a bias for

شكل رقم (١١) يمثل نموذجا لاحتواء النص على نصوص متعددة اللغات أو متعددة الرموز أو الحروف.

## المعيار الثاني:

التعرف على علامات الضبط بالشكل (علامات التشكيل) والتفرقة بينها وبين نقط الإعجام:

تتميز اللغة العربية بالتعبير عن الأصوات (vowels) بحركات الضبط، وتأتي هذه الحركات مقابلة للصوائت في اللغات الأخرى (e-o-i-u-a)، وهكذا يعتور الكلمة العربية الكثير من الزوائد الشكلية من تنقيط الحروف وحركات الضبط بالشكل.

| ـَ | ـً | ـُ | ـٌ | ـِ | ـٍ | ـْ |
|---|---|---|---|---|---|---|
| فتحة | تنوين بالفتحة | ضمة | تنوين بالضمة | كسرة | تنوين بالكسرة | سكون |

| ـَّ | ـًّ | ـُّ | ـٌّ | ـِّ | ـٍّ |
|---|---|---|---|---|---|
| شدة وفتحة | شدة وتنوين بالفتحة | شدة وضمة | شدة وتنوين بالضمة | شدة وكسرة | شدة وتنوين بالكسرة |

جدول رقم (٧) لعلامات الضبط بالشكل في العربية

والحق أن اللغة العربية المعاصرة لا تستخدم علامات الضبط بالشكل في النصوص المعاصرة، ومن ثم يمكن مراعاة هذا المعيار في النصوص الكلاسيكية وحسب، وتجاهله في النصوص المعاصرة.

## المعيار الثالث:

التعرف على الأرقام العربية بصورتيها؛ الهندية ١ – ٢ – ٣ – والعربية –3 – 2 – 1 ...

## المعيار الرابع:

التعرف على علامات الترقيم ( ، – . – ؛ – ؟ – ! ). ومن المهم هنا الانتباه إلى أن هذه العلامات قد تتصل بالكلمة تارة وقد لا تتصل بها تارة أخرى، ومن ثم ينبغي معالجة هذه العلامات في كلتا صورتيها.

## المعيار الخامس:

التعرف على الرموز والاختصارات المختلفة (@ – # – $ – ٪ – [...] ).

## المعيار السادس:

التعرف على الصور والأشكال الرسومية، ونقلها مرة أخرى إلى الملف الرقمي في مكانها من الملف الأصلي.

## المعيار السابع:

مدى استخدام المعارف اللغوية في عملية التعرف الآلي على الحروف (من خلال مراعاة خصائص الحروف من حيث الشكل والترتيب والضبط بالشكل)

فللحروف العربية خصائص من حيث الشكل والموضع من الكلمة، فهناك حروف تأتي في أول الكلمة فقط، وهناك حروف تأتي في أوسطها ومنتصفها فقط ، وهناك حروف متطرفة تأتي في أواخر الكلمة فقط (الحروف الاستهلالية – الحروف المتوسطة– الحروف المتطرفة)، بل يتغير رسم الحرف باختلاف ترتيبه مع الحروف الأخرى بداية أو توسطا أو نهاية.

**ومثال ذلك:**

**الألف والهمزة**

تختلف أشكال كتابة الألف والهمزة – على سبيل الحصر – على النحو التالي:

**{أ – إ – ا – ء – آ – ؤ – ئ – ى}**

فجميع هذه الصور تأتي في بداية الكلمة ماعدا

**{ء – ؤ – ئ – ى}**

وجميعها أيضا تأتي في وسط الكلمة ماعدا

**{ى}**

وجميعها كلها تأتي في نهاية الكلمة ماعدا

**{إ}**

وقد تجتمع أكثر من همزة في بداية الكلمة ولكن على النحو الآتي فقط :

**{أأ – أئ – أؤ – أ – آل}**

ومن ذلك أيضا مراعاة قواعد الرسم الإملائي في العربية، ومنها:

توالي الأمثال في العربية [1]:

تهرب العربية من توالي الأمثال إذا توالى حرفان أحيانًا، وتهرب منه بشكل قاطع إذا توالت ثلاثة أحرف.

ويمكن حصر الحروف التي لا تتوالى مرتين ، ومن ثم تخطئتها إذا جاءت مكررة، أما توالي ٣ حروف فممنوع في العربية منعا قاطعا مثل( لللـ ) في (لllيمون) وصوابها (لليمون)، ومن ثم يمكن حذف الحرف المكرر الثالث تلقائيًّا أو استبداله بما يشبهه.

ومن ذلك أيضا – عند التعرف على النصوص المضبوطة بالشكل- مراعاة قواعد الضبط والتشكيل:

أ- الفتحة والكسرة والضمة والسكون، تأتي في جميع مواضع الحروف من الكلمة ( الحرف الأول والثاني والثالث، ...

ب- التنوين بالفتحة والكسرة والضمة يأتي في نهايات الكلمة فقط (على الحرف الأخير) ولا يأتي التنوين في بداية الكلمة (على الحرف الأول) أو في وسطها (وقد يأتي التنوين قبل الآخر في حال التنوين بالفتحة فقط في إحدى صور كتابة التنوين بالفتحة).

ج- التنوين بالفتحة يزيد حرفا هو الألف على الكلمة غير المنتهية بتاء مربوطة أو ياء.

د- السكون لا يأتي في الحرف الأول من الكلمة العربية والقاعدة اللغوية في ذلك مشهورة: لا يُبدأ بساكن .

هـ - لا يتوالى ساكنان في العربية (ــْــْ ) [2].

و- الشدة لا تأتي في الحرف الأول من الكلمة العربية؛ ذلك أن الشدة مكوّنة من حرفين: حرف ساكن+ حرف متحرك.

ز- ألف الوصل لا تضبط بالشكل.

---

١ - راجع: بحوث ومقالات في اللغة ، د. رمضان عبد التواب، مكتبة الخانجي بالقاهرة، ط .٣ ، ١٤١٥هـ -١٩٩٥م ، ص ٢٧.

٢ - راجع: النحو الوافي. د. عباس حسن. دار المعارف. ط٣. مصر، القاهرة، ج٢. ص٤٦٦ ـ ج٣.ص٢٠٠.

### المعيار الثامن:

وجود تقنية لغوية متقدمة تمكنها من التصحيح التلقائي لأخطاء القراءة بالاعتماد على مدقق إملائي يعتمد على محلل صرفي أو من خلال معجم لغوي مزود بالتطبيق تكون بمثابة المرجع في التصحيح التلقائي.

### المعيار التاسع:

مدى الاستفادة من تصحيحات المستخدم وتدريب برامج التعرف الآلي على التعلم من أخطائه السابقة، ومن ثم تكوين معجم مفردات خاص بالمستخدم، يضمن التصحيح وفقًا للحقل اللغوي الذي عادة ما يكتب المستخدم فيه، أو وفقًا لمفردات المستخدم التي يستخدمها بكثرة على هذا التطبيق.



# الفصل الثاني
# رؤى مطروحة لتطوير بعض التقنيات اللغوية

## المبحث الأول
## تطوير تقنيات التدقيق الإملائي والنحوي لتفعيل دورها في تعليم العربية

### تمهيد:

يهدف هذا المبحث إلى إبراز دور تقنيات اللغة العربية الحاسوبية في التعليم الإلكتروني، من خلال اتخاذ المدققات الإملائية والنحوية نموذجًا ومثالاً؛ إذ إنّ من الممكن استخدام هذه التقنيات الإلكترونية في تعليم اللغة العربية لمستخدميها بطريقة غير مباشرة، فعادة ما يعتمد المستخدم العربي على هذه التطبيقات (المدققات الإملائية والنحوية) في الوصول إلى عربية صحيحة لنصوصه في برامج معالجة النصوص العربية، وينتهي دور التقنية بعد تحقيق هدفها التطبيقي، لكن المبحث الذي بين أيدينا يهدف إلى أن يمتد عَمَلُ هذه التقنيات اللغوية إلى الإسهام في تعليم مستخدميها قواعد العربية كذلك، إضافةً إلى القيام بدورها الأساسي في التدقيق الإملائي والنحوي.

يتناول المبحث التعريف بتطبيقات التدقيق الإملائي والنحوي proofing tools، وأشهر المدققات الإملائية والنحوية، ثم يعرض للوظيفة الأساسية لهذه التطبيقات وكيفية استخدام هذه التقنيات اللغوية في تعليم العربية للمستخدمين العرب بطريقة إلكترونية، من خلال عدد من الأخطاء الإملائية والنحوية الشائعة التي تعالجها تلك التطبيقات، مثل: الخلط بين أشكال همزتي القطع والوصل في أول الكلمة، والخلط بين التاء المربوطة (ـة) والهاء (ـه) في نهاية الكلمة العربية، والخلط بين أشكال الياء

في نهاية الكلمة العربية، والخلط بين علامات الإعراب بالحروف (المثنى - جمع المذكر السالم - الفعل المضارع - الأفعال الخمسة - الفعل الأمر) ، وعدم المطابقة بين عنصري الإسناد، ... وغير ذلك من الأخطاء الإملائية والنحوية.

أما الآليّة الجديدة المقترحة في هذ المبحث فنعرض من خلالها للإضافات التعليمية التي تسهم في تطوير تعلم العربية إلكترونيا، مثل: توصيف الخطأ وتحليله، والباب النحوي للمسألة وعلاقتها بشجرة العلاقات النحوية، والاستماع إلى الشاهد التراثي في المسألة وعرضه مع تلوين الكلمة محل الشاهد، وتوثيق القاعدة من مراجع العربية، والربط بين الأخطاء الشائعة عند المستخدم أو عند المستخدمين العرب جميعا، والتنبيه عند تكرار نوع الخطأ من المستخدم بشكل خاص أو عام (خطأ شائع).

## مقدمة

يقترح المبحث فكرة آلية لغوية هدفها تقريب العربية من أهلها، واستخدام التقنيات الحديثة في تعلم العربية وقواعدها، وهي لا تتعامل مع المتعلم بطريقة الدروس التقليدية درسًا درسًا، وإنما من خلال التطبيق العملي على النصوص، ومن خلال معالجة أخطاء المستخدم نفسه، وتحليل هذه الأخطاء وردّها إلى قضاياها الأساسية وأصولها الكلية في العربية نحوًا وصرفًا.

والآلية المقترحة أداةٌ لغوية مدمجة (Plugin) تعتمد على مخرجات برامج التدقيق اللغوي، وعلى قاعدة بيانات تحليلية تقارن الآليةُ من خلالها الحالةَ التي عليها الكلمة أو الجملة بالحالةِ المقترحة من برامج التدقيق اللغوي، ومن ثم تتمكن الآلية من تقديم التأصيل العلمي للقضية محل الإشكال إملائيًّا أو نحويًّا أو تركيبيًّا. ولا تكاد الدراسة تتعرض لوجهة النظر الحاسوبية التقنية، وإنما تعرض للجانب اللغوي فقط؛ لتخصص الباحث فيه، تاركة الجانب التقني البرمجي إلى بحث آخر.

أما معوقات هذه الآلية فتتلخص في ضعف بعض مخرجات برامج التدقيق الإملائي والنحوي وعدم استقصائها كثيرًا من المشكلات اللغوية، كمشكلات اللبس اللغوي بنيةً وتركيبًا، لكنها في الوقت نفسه تصلح لأن تكون أساسًا يبنى في تعليم العربية بطريقة إلكترونية جذابة، مع دعوة الباحثين إلى تركيز جهودهم لحل مشكلات برامج

التدقيق الإملائي والنحوي، ومن أهمها مشكلة اللبس في البنية والتركيب، فكثير من مستخدمي تطبيقات التدقيق الإملائي -على سبيل المثال- يضيقون ذرعًا بهذه التطبيقات، وذلك بسبب فشلها في اكتشاف الكثير من الأخطاء الإملائية في سياقات معينة، فالكلمة «اضرب» في سياق جملة زمنها المضارع (دعوني اضرب لكم مثلا)؛ تُعدّ خطأً إملائيًّا يلزم تصويبه إلى الكلمة «أضرب» بهمزة قطع، والكلمة «السنة» في سياق التركيب «السنة النيران» تعدّ خطأً إملائيًّا، لكن المدقّق يفشل في التعرف على مثل هذه الأخطاء، ومن ثم في اقتراح تصويبها، ويرجع السبب في ذلك إلى اللبس الصرفي، فالمدقق الإملائي يتعامل مع الكلمة المفردة بلا نظرٍ إلى سياقها، فهو يتعامل مع الكلمة في مستواها الصرفي فقط، ومن ثمّ فإن أية احتمالات صرفية سليمة أو صحيحة تحمل المدقق على عدم تخطئة الكلمة، على الرغم من كونها خطأً سياقيًّا[1]، ففي الجملة: «شعب من فصيلة الصبار» لا يمكن للمدقق أن يكتشف الخطأ في كلمة «شعب» ومن ثم لا يقترح الصواب»عشب»، وذلك لأن كلمة «شعب» صحيحة صرفيًّا ولكنها خطأ في هذا السياق، وليس من شأن المدققات الإملائية الحالية أن تنظر إلى السياق، بل إنها تكتفي بالتحليل الصرفي للكلمة. ومثل ذلك التراكيب، ومنها: «بنات نادر»، وصوابها «نبات نادر» ، و»تعقدت الجرحى» ، وصوابها «تفقدت الجرحى»، فهذه الأخطاء جميعها تحتمل وجهًا من الصواب منفردة(دون سياق)، ولكنها في سياقاتها تحتاج إلى التصحيح، ومن هنا يتضح بجلاء مدى انعكاس ظاهرة اللبس الصرفي على عمل المدقق الإملائي.

---

١ - راجع: مشكلات التحليل النصي للمحتوى العربي على شبكة الإنترنت، د. سلوى حمادة، عمرو جمعة، كتاب مؤتمر هندسة اللغة الثامن، ٢٠٠٨، جامعة عين شمس، كلية الهندسة.

## الدراسات السابقة:

المدققان الإملائي والنحوي آليتان برمجيتان تتعاملان مع النصوص العربية ( في الكلمة والتركيب معًا)، وتتلخص وظيفتاهما في مراجعة صحة الكلمة إملاءً وإعرابًا وصحة الجملة تركيبًا، ومن ثم اقتراح بدائل صحيحة للكلمة غير الصحيحة إملاءً أو نحوًا، أو اقتراح تعديل للجملة غير الصحيحة تركيبًا؛ لأجل الوصول إلى الصواب النحوي أو التركيبي. فالمدقق الإملائي في برامج مايكروسوفت لا تقدم تعليلاً تعليميًّا للصواب المقترح من قبلها، وإنما تكتفي باقتراح البدائل، أما المدقق النحوي الخاص بتطبيقات مايكروسوفت فيضيف إلى اكتشاف الخطأ وتصحيحه بعدًا تعليليًّا تعليميًّا وهو الإشارة إلى نص القاعدة التي بنى المدقق عليها اكتشافَه الخطأ والتنبيه إليه، لكن هذه الإشارة التعليمية لا تعدو أن تكون جملة واحدة - مثل: «لابد أن يتفق النعت والمنعوت في العدد والنوع» - وهذه الجملة تناسب الغرض التطبيقي لا التعليمي من البرنامج. أما الآلية التعليمية المقترحة فتضيف إلى الهدف التطبيقي اللغوي بعدًا تعليميًّا آخر، على غرار البعد التطبيقي، يشمل مع كل اقتراح للتصويب: البحثَ على الإنترنت، وتحليلَ الخطأ، والقاعدة، والمرجع المعتمد في ذلك (مع رابطه على الإنترنت)، والشاهد الفصيح (مع إمكان الاستماع إليه)، وتحديد نوع الخطا إن كان من الأخطاء الشائعة مثلاً، بالإضافة إلى تحديد مدى تكرار الخطأ لدى المستخدم.

## التعريف بتطبيقات التدقيق الإملائي والنحوي proofing tools:

إن من أشهر المدققات الإملائية مدقق برنامج مايكروسوفت وورد، الذي طرح لأول مرة مع «وورد» في الإصدار ٦,٠ منه عام ١٩٩٢، ولم يكن من إنتاج شركة مايكروسوفت ذاتها، بل اشترته من شركة كولتك، التي يقع مقرها الرئيس في القاهرة، وهي أيضاً الشركة التي طورت المدقق النحوي لمايكروسوفت وورد، الذي طرح أول مرة مع الإصدار عام ٢٠٠٠م[1].

وهو سهل الاستخدام، ويطرح بدائل عديدة لتصحيح الكلمة لغير المتخصصين،

---

١ - راجع: موقع شركة كولتك - الشرق الأوسط لتقنيات معالجة اللغات الطبيعية على الإنترنت، الرابط التالي: http: //www.coltec.net /default.aspx?tabid=224

ويميز بين الخطأ الإملائي والخطأ النحوي عن طريق لون الخط المتموج الموجود تحت الكلمة الخطأ؛ فاللون الأحمر يعني خطأً إملائيًّا، واللون الأخضر يعني خطأً نحويًّا أو تركيبيًّا. وإذا ما نقر المستخدم بزر الفأرة الأيمن على الخط المتموج، فستظهر له نافذة صغيرة تحوي كلمات أخرى مقترحة لتصحيح الخطأ الموجود في حالة الخطأ الإملائي، أما في حالة الخطأ النحوي فتظهر النافذة شرحًا مبسطًا للخطأ النحوي مع الكلمة المقترحة لتصحيحه. ففي جملة «لا تكن من المهملون» - على سبيل المثال - يضع البرنامج خطًّا أحمر تحت كلمة «المهملون» علامة على وجود الخطأ، ثم يخرج البرنامج رسالة مضمونها: «الاسم مجرور بحرف الجر، تُستبدَل كلمة «المهملين» بكلمة «المهملون» لتصحيح الخطأ».

ومن أشهر المدققات كذلك مدقق شركة صخر، واسمه «المصحح الآلي» ومن الممكن دمجه في برنامج مايكروسوفت وورد، أو استخدامه مع بعض البرامج الأخرى، ويكتشف المصحح الآلي الأخطاء التي تقع عند الكتابة بالعربية، ويقترح مجوعة من البدائل الصحيحة، ثم يبدأ في محاولة اكتشاف الأخطاء النحوية.

ويتميز المصحح عند تصحيحه النصوص العربية عن غيره من برامج التصحيح بعدد من المميزات وهي: التصحيح التلقائي، وتقليل عدد الاقتراحات، ومراعاة السياق، والتعامل مع المصطلحات، والتعامل مع التشكيل[1].

ويتعامل المدقق الإملائي والنحوي -حسب تسميتها- مع نوعين من الأخطاء في النصوص العربية، فالمدقق الإملائي كما يبدو من اسمه يتعامل مع الأخطاء الإملائية، والمدقق النحوي يتعامل مع الأخطاء النحوية.

## استخدام التقنيات اللغوية في تعليم العربية إلكترونيًّا:

الآلية المقترحة تضيف إلى التصحيح الآلي بُعدًا تعليميًّا تطبيقيًّا، إذ لا يكفي المستخدم حينئذ بتصحيح الخطأ فحسب، بل يستطيع من خلالها أن يتعلم قاعدة جديدة من قواعد العربية، وأن يدعم التصحيح المقترح بتحليل الخطأ الموجود وأساس تصويبه، والبحث

---

١ - راجع: موقع شركة صخر لبرامج الحاسب، الرابط :
http: //www.sakhr.com /arabicresources.aspx.

عن الكلمة واستخداماتها على الإنترنت، كما يمكنه أن يطلع على توثيق ذلك من المراجع اللغوية العربية، ومن ثم معرفة الباب النحوي الذي تنتمي إليه المسألة، ومن ثم يدرك علاقتها بشجرة النحو العربي، ويمكنه كذلك أن يطلع على شواهد مماثلة من نصوص التراث العربي في القرآن الكريم وأقوال العرب وأشعارهم، وأن يستمع إليها كذلك بصوت المشاهير، ذلك كله من خلال روابط متصلة بصفحة على الإنترنت، أو قاعدة بيانات لغوية.

ويمكن للآلية الربط بين الأخطاء الواحدة والمتشابهة، والتعرف على تكرار هذا الخطأ عند المستخدم الواحد، وتحديد نوع الخطأ إن كان من الأخطاء الشائعة عند المستخدمين العرب. كما يمكن للآلية أن تنبّه المستخدم أثناء الكتابة لتجنب الوقوع في الخطأ نفسه في حال تكرّر نوعُه، من خلال ذاكرة أخطاء المستخدم الخاصة أو ذاكرة الأخطاء الشائعة العامة.

**والجدول الآتي يلخص المعلومات الجديدة التي تضيفها الآلية بوصفها وسيلة تعليمية إلكترونية.**

| إضافات آلية التعليم الإلكتروني | |
|---|---|
| الخطأ | توصيف الخطأ وتحليله |
| البحث | البحث في جوجل |
| المزيد من المعلومات | الباب النحوي للمسألة وعلاقتها بشجرة العلاقات النحوية |
| سماع | الاستماع إلى الشاهد التراثي في المسألة وعرضه مع تلوين الكلمة محل الشاهد |
| التوثيق | توثيق القاعدة من المراجع العربية |
| التكرار | الربط بين الأخطاء الشائعة من المستخدم بخاصة أو من المستخدمين العرب بعامة |
| التنبية | التنبيه عند تكرار نوع الخطأ من المستخدم بشكل خاص أو عام (خطأ شائع). |

جدول رقم (١٩) لإضافات آلية التعليم الإلكتروني

**وسنعرض فيما يأتي عددًا من الأخطاء الإملائية والنحوية الشائعة، ونشرح كيفية الإفادة من تصحيح هذه الأخطاء في تعليم العربية للمستخدمين العرب بطريقة إلكترونية.**

## أولا المدقق الإملائي:

**من الأخطاء الإملائية الشائعة في النصوص العربية(١):**

**- الخلط بين أشكال همزتي القطع والوصل في أول الكلمة[٢]:**

فرسم الهمزة في الكلمة «إنتقد» - على سبيل المثال - خطأ إملائي، يكتشفه المدقق الإملائي تلقائيًّا وينبه إليه، ويقترح بدائله، ثم يأتي عمل آلية التعليم الإلكتروني بعد ذلك في عرض توصيف الخطأ وتحليله، وإمكان البحث عن الكلمة في جوجل، وربط المسألة بالباب النحوي لها، وبيان علاقتها بشجرة العلاقات النحوية، والاستماع إلى الشاهد التراثي في المسألة وعرضه، مع تلوين الكلمة محل الشاهد، وتوثيق القاعدة من المراجع العربية، والربط بين الأخطاء الشائعة عند المستخدم نفسه أو المستخدمين العرب جميعا، ثم التنبيه عند تكرار نوع الخطأ من المستخدم بشكل خاص أو عام (خطأ شائع).

شكل رقم (١٢) لمعالجة المدقق الإملائي كلمةَ "إنتقد" بهدف تعليمي.

١ - راجع: أحكام الإملاء في اللغة العربية وقواعده في:
الهمزة في اللغة العربية، مصطفى التوني، دراسة لغوية، القاهرة ١٩٩٠م.
دراسة في قواعد الإملاء، عبد الجواد الطيب،دار الأوزاعي، ط. ثانية، بيروت ١٤٠٦هـ/ ١٩٨٦م.
قواعد الإملاء، عبد السلام هارون، دار إيلاف الدولية، الكويت، ط. أولى، ١٤٢٥هـ/ ٢٠٠٤م.

٢ - راجع: الإملاء والترقيم في الكتابة العربية، عبد العليم إبراهيم، ص ٣٧-٣٩.

**- الخلط بين التاء المربوطة (ـة) والهاء (ـه) في نهاية الكلمة العربية:**

فكتابة كلمة «المتحده» بالهاء لا بالتاء - على سبيل المثال - خطأ إملائي شائع، يكتشفه المدقق الإملائي تلقائيًّا وينبه إليه ويقترح بدائله، ثم يأتي عمل آلية التعليم الإلكتروني كما يوضح الشكل الآتي:

شكل رقم (١٣) لمعالجة المدقق الإملائي كلمةَ "المتحده" بهدف تعليمي.

**- الخلط بين أشكال الياء في نهاية الكلمة العربية:**

فالياء في نهاية الكلمة العربية إما أن تكون ألفا مقصورة (ى) أو ياء منقوصة (ي)، فكلمة «منطقي» بمنقوطتين في الكتابة الشامية يمكن أن تكتب خطأ هكذا: «منطقى» بياء غير منقوطة.

**- الخطأ في رسم الهمزات المتوسطة والمتطرفة (ء، ئ، ؤ، أ، آ) في الكلمة العربية:**

فكلمة «شيء» بهمزة على السطر يمكن أن تكتب خطأً هكذا: «شيئ» بهمزة على نبرة[1].

**- الخطأ في وصل كلمتين دون الفصل بينهما بمسافة:**

فالكلمتان «يا الله» يمكن أن تكتبا خطأ هكذا «ياالله» دون مسافة بينهما.

**- وجود نقص في حروف الكلمة العربية:**

فكلمة «المحامي» يمكن أن تكتب خطأً هكذا: «المحام».

---

١ - راجع: الإملاء والترقيم في الكتابة العربية، عبد العليم إبراهيم، ص ٥٦-٥٧.

**- وجود تكرار أو زيادة في الكلمة مما تمجّها العربية:**

فكلمة «مشكور» تكتب عادة - لاسيما في وسائل الإعلام الاجتماعي - خطأً هكذا: «مشكووووور» للدلالة على المبالغة في الأمر.

**- وجود قلب مكاني بين حروف الكلمة العربية لعوامل لهجية:**

فكلمة «يحفر» يمكن أن تكتب خطأ هكذا: «يفحر».

**- استبدال بعض الحروف بحروف الكلمة لعوامل لهجية:**

ومثاله إبدال البعض حرف الجيم «ياء» في كلمة «رجال»، فيكتبونها «ريال» تأثرا باللهجة.

**- الخلط بين الحروف المتجاورة في لوحة مفاتيح الحاسوب:**

فالبعض يخطئ في كتابة بعض الحروف نتيجة لسرعته في الكتابة على لوحة المفاتيح، فكلمة «فاروق» على سبيل المثال تكتب خطأً هكذا: «فاروف» لتجاور حرفي «الفاء» و»القاف» في لوحة المفاتيح، ومن أمثلته أيضًا الحروف التي تكتب بطريقة مركبة على لوحة المفاتيح باستخدام مفتاح «shift"، فكلمة "إيمان" يمكن أن تكتب خطأ هكذا: "غيمان"، فالهمزة المكسورة تكتب حاسوبيًّا بالضغط على "غ+shift"، وفي حال عدم الضغط على "shift" تكتب "غَيْنًا" على الفور.

هذه بعض الأخطاء الإملائية الشائعة، التي يقع المستخدم العربي فيها أثناء استخدامه برامجَ النصوص، وتستطيع برامج التدقيق الإملائي التعرف على هذه الأخطاء وتصحيحها، لكن بِلا تعليم للمستخدم وتدريب له على القاعدة التي تحكم بالصواب أو بالخطأ في المسألة. ويقترح المبحث - كما عرضنا - أن تضاف إلى برامج التدقيق الإملائي آلية تضم بعض القواعد اللغوية المبسطة التي تشرح الخطأ وتعرض أساس تصحيحه؛ تجنّبًا لوقوع المستخدم في الخطأ مرة أخرى، ومن ثم تطوير طرق تعليم اللغة العربية إلكترونيا من خلال التقنيات اللغوية.

## ثانيا المدقق النحوي:

**التعريف بالمدقق النحوي Arabic Grammar Checker:**

هو آلية برمجية تتعامل مع النصوص العربية ( الكلمة والتركيب معًا)، وتتلخص وظيفة المدقق النحوي في مراجعة صحة الكلمة إعرابًا وصحة الجملة تركيبًا، ثم اقتراح بدائل صحيحة للكلمة غير الصحيحة نحوًا، أو اقتراح تعديل للجملة غير الصحيحة تركيبًا؛ لأجل الوصول إلى الصواب النحوي أو التركيبي.

**كيفية عمل المدقق النحوي:**

إن البحث لا يقدم مدقّقًا نحويًّا، وإنما يقدم برمجةً يمكن إضافتها إلى المدقق النحوي ليكون أداة تعليمية يمكن من خلالها تعلم قواعد العربية بطريقة إلكترونية تطبيقية عملية، ومن ثم فالآلية الجديدة تعتمد على المدقق النحوي في اكتشاف الخطأ النحوي، وهو الذي يعتمد في كثير من عمله على المدقق الإملائي وأداتيه؛ أداة التحقق Verification Tool، وأداة المقترحات Suggestions Tool[1]، كما يعتمد كذلك على بعض الآليات اللغوية الأخرى المضمنة في المدقق النحوي ومنها – على سبيل المثال – Part Of Speech Tagger ، حيث يتعرف من خلال هذه الأداة على نوع الكلمة من حيث كونها اسمًا أو فعلًا أو حرفًا، ثم يحدّد سماتها التركيبية والدلالية، ويدقق الجمل من حيث النحو والتركيب.

والجدير بالذكر هنا أن المدقق النحوي الخاص بتطبيقات مايكروسوفت يضيف إلى اكتشاف الخطأ وتصحيحه بُعدًا تعليليًّا تعليميًّا، وهو الإشارة إلى نص القاعدة التي بنى المدقق عليها اكتشافه الخطأ والتنبيه إليه، لكن هذه الإشارة التعليمية لا تعدو أن تكون جملة واحدة، مثل: "لابد أن يتفق النعت والمنعوت في العدد والنوع"، وهذه الجملة تناسب الغرض التطبيقي لا التعليمي من البرنامج.

أما الآلية التعليمية المقترحة فتضيف إلى الهدف التطبيقي اللغوي بُعدًا تعليميًّا آخر، على غرار البعد التطبيقي، يشمل مع كل اقتراح للتصويب: البحثَ على الإنترنت، وتحليلَ الخطأ، والقاعدة، والمرجع المعتمد في ذلك (مع رابطه على الإنترنت)، والشاهد

---

١ – راجع: موقع شركة كولتك – الشرق الأوسط لتقنيات معالجة اللغات الطبيعية على الإنترنت www.coltec.com.

الفصيح (مع إمكان الاستماع إليه)، وتحديد نوع الخطأ إم كان من الأخطاء الشائعة، إضافة إلى تحديد مدى تكرار الخطأ لدى المستخدم.

**وسنعرض فيما يأتي أمثلة لبعض الأخطاء النحوية الشائعة في النصوص العربية:**

### ١- الخلط بين علامات الإعراب بالحروف (المثنى - جمع المذكر السالم - الفعل المضارع -الأفعال الخمسة- الفعل الأمر)

عدم المطابقة بين عنصري الإسناد.

**الخطأ في كتابة العدد بالحروف في اللغة العربية، وذلك على التفصيل الآتي:**

الخلط بين علامات الإعراب بالحروف (المثنى - جمع المذكر السالم - الفعل المضارع- الأفعال الخمسة - الفعل الأمر).

الخطأ في علامة الإعراب من الأخطاء الشائعة لدى المستخدم العربي، لكن هذا الخطأ لا يظهر إلا في حالة الإعراب بالحروف، فالعربية المعاصرة المكتوبة تتجاهل حركات الإعراب من الفتحة والضمة والكسرة إضافة إلى عدم الحركة أو السكون، ومن ثم فالخطأ في أي من حركات الإعراب لن يظهر؛ لعدم ظهور علامات الإعراب على الكلمات في العربية المعاصرة المكتوبة، وذلك على العكس من علامات الإعراب بالحروف (الألف والياء، والياء والواو، وحذف حرف العلة، وثبوت النون، ...) [١] التي تظهر جليًّا لكون علامات الإعراب فيها جزءًا لا يتجزأ من الكلمة، فهي لواحق تلحق بساق الكلمة.

وتنبني فكرة عمل المدقق النحوي في بعض أجزائها على السمات النحوية والتركيبية لسوابق الكلمة الواحدة ولواحقها؛ فالكلمة صرفيًّا تتكون من ساق Stem وزوائد Affix (هي السوابق واللواحق)، والكلمة الصحيحة إملاءً هي الكلمة الصحيحة من حيث سلامة ساقها ومن حيث قبول هذه الساق لتلك السوابق واللواحق الخاصة بها، ثم إن لهذه السوابق واللواحق سماتٍ نحوية وتركيبية تمكن المدقق النحوي من اكتشاف الأخطاء النحوية .

وأوضح الأمثلة على ذلك (من سوابق الأسماء) حروف الجر المتصلة بالكلمة؛ الكاف (كَـ) ، والباء (بِـ) ، واللام (لِـ). فهذه الحروف تجر الأسماء التي تدخل عليها، وتختلف

---

١ - راجع: النحو الأساسي، د. أحمد مختار، د. مصطفى النحاس، د. محمد حماسة، دار السلاسل، الكويت، ١٩٩٤، ط ٤، ص٦٣- ٦٤.

علامات الجر باختلاف هذه الأسماء؛ ما بين الكسرة في حال المفرد، وما يجمع بالألف والتاء، والفتحة في حال الممنوع من الصرف، والياء في حالتي المثنى وجمع المذكر السالم. لكن أثرها يتجلى - كما سبق أن ذكرنا - في المستوى الكتابي للعربية المعاصرة في المثنى وجمع المذكر السالم منها دون غيرهما لكون الإعراب فيهما بالحروف دون الحركات التي تتجاهلها العربية المعاصرة المكتوبة حيث تكتب العربية المعاصرة دون علامات الضبط.

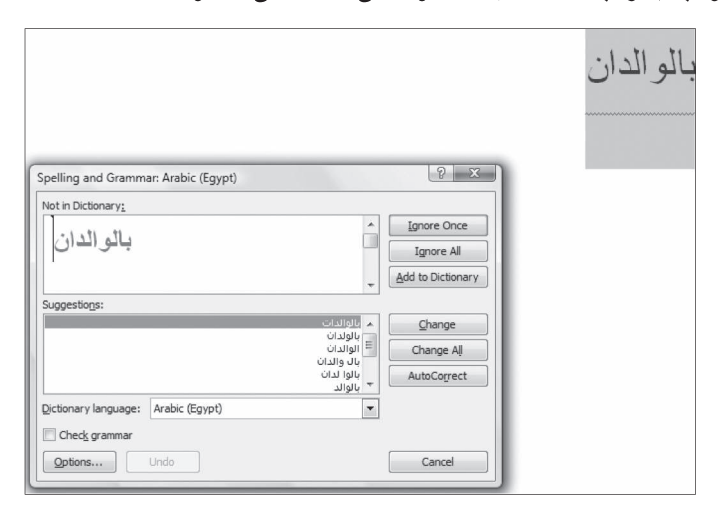
فعند معالجة المدقق النحوي لكلمة «بالوالدان» - على سبيل المثال - سيقوم المحلل الصرفي المضمن في المدقق الإملائي بتحليلها على النحو التالي:

«بالـ» (سابقة) = «بـ» (حرف جر) + «ألـ» (التعريفية)

"والد" = (اسم)

"ـان" (لاحقة) = علامة رفع المثنى

لكن السابقة («بالـ» = حرف جر + ألـ التعريفية) لا تتوافق نحويًّا مع اللاحقة («ـان» = علامة رفع المثنى)، إذ كيف يجتمع حرف الجر مع علامة الرفع في كلمة واحدة، ومن ثم يبدأ النظام في تخطئة الكلمة، واقتراح احتمالات الصواب لها، وعلى رأسها تصويب إعراب الكلمة إلى «بالوالدين» بدلا من «بالوالدان».




شكل رقم (١٤) لمعالجة المدقق النحوي كلمةَ "بالوالدان".

ولا يعتمد المدقق النحوي - في اكتشافه الأخطاءَ - على السمات النحوية والتركيبية للسوابق واللواحق، وإنما يطبق الأمر ذاته على الأدوات وحروف المعاني، فحروف

الجر المنفصلة (من، إلى، في، عن، على) مِثل حروف الجر المتصلة (ل، ب، ك) تتوافق – على سبيل المثال – مع بعض اللواحق المتصلة بالاسم المجرور، ففي التركيب "إلى الحديقتان"، لا يمكن أن تجتمع اللاحقة "ـان" في كلمة "الحديقتان" التي تدل على الرفع مع حرف الجر "إلى" صاحب تأثير الجر في الأسماء الواقعة بعده، ومن ثم يستشكل المدقق النحوي الكلمة ويخطّئها، ويقترح تصويبها "الحديقتين".

ومنه أيضا التركيب "لن يذهبون"، حيث يستحيل أن تجتمع اللاحقة "ـون" في كلمة "يذهبون" والأداة "لن" التي تنصب الفعل المضارع، فعلامة نصبه هنا حذف النون، ومن ثم يستشكل المدقق النحوي الكلمة، ويقترح تصويبها "يذهبوا".

وخلاصة القول: إن للزوائد من السوابق واللواحق وظيفتين؛ تصريفية وتركيبية، وتحاول تطبيقات التدقيق الإملائي والنحوي توظيف هاتين الوظيفتين التصريفية والتركيبية في عملية التدقيق الآلي.

هنا انتهى البعد التطبيقي للمدقق النحوي، إذ اقتصر عمله على اكتشاف الخطأ والتنبيه إليه، ويأتي بعد ذلك دور الآلية التعليمية في التعليل والتوضيح لكل اقتراح تصويبي من قبل المدقق النحوي، ففي المثال: "بالوالدان" تتعامل الآلية على النحو التالي:



شكل رقم (١٥) لمعالجة المدقق النحوي كلمةَ "بالوالدان" بهدف تعليمي.

## ٢- عدم المطابقة بين عنصري الإسناد أو التركيب:

بالإضافة إلى اعتماد المدقق النحوي على السوابق واللواحق في اكتشاف الأخطاء الإعرابية فهو يعتمد عليها كذلك في اكتشاف وتصحيح العديد من الأخطاء التركيبية، مثل المطابقة في العدد، والتعريف والتنكير، والنوع، والإعراب بين كل من المسند والمسند إليه (الفعل والفاعل أو الفعل ونائب الفاعل- المبتدأ والخبر)، والصفة والموصوف، والضمير ومرجعه، والبدل والمبدل منه، والحال وصاحب الحال، واسم الإشارة والمشار إليه.

- ففي هذه الجملة: "الجنود يدافعان عن الوطن" خطأ نحوي يتلخص في عدم التطابق بين المسند والمسند إليه في العدد.

شكل رقم (١٦) لمعالجة المدقق النحوي كلمةَ "يدافعان".

فالمسند إليه "الجنود" في الجملة السابقة جمع تكسير، والمسند "يدافعان" مسند إلى ألف الاثنين، وهو خطأ بيِّنٌ في العربية، صوابه أن يتفق ضمير الفاعل مع العائد عليه في العدد والنوع[1].

والمدقق هنا - كما سبق أن أوضحنا - يعرض للقاعدة التي تحكم بتصويبه الخطأَ النحوي، غير أنه لا يضيف أي بعد تعليمي للمسألة، وهو ما يمكن معالجته على النحو الآتي:

١ - راجع: النحو الأساسي، د. أحمد مختار، د. مصطفى النحاس، د. محمد حماسة، ص ٤٣٢-٤٣٣.




شكل رقم (١٧) لمعالجة المدقق النحوي كلمةَ "يدافعان" بهدف تعليمي.

- **وفي الجملة الآتية**: "الرجال الكريم يساعدون الفقراء" خطأ بيِّن، فكلمة "الكريم" نعت مفرد يجب أن يتفق مع المنعوت الجمع "الرجال" في العدد، وصوابه "الرجال الكرماء أو الكريمون ..."




شكل رقم (١٨) لمعالجة المدقق النحوي التركيبَ "الرجال الكريم".

ومثله كذلك: التطابق بين اسم الإشارة والمشار إليه في العدد، كما في الجملة: "ذهبنا إلى هؤلاء المعلم"، وصوابه: "ذهبنا إلى هذا المعلم". ومن التطابق بين عنصري الإسناد

التطابق بين الفعل والفاعل في النوع داخل الجملة: «كرهت الولد عصير البرتقال» وصوابه: «كره الولد العصير».

### ٣- الخطأ في كتابة العدد بالحروف في اللغة العربية:

لكتابة العدد بالحروف العربية قواعد محددة من حيث تذكير العدد وتأنيثه حسب المعدود، وإفراد المعدود وجمعه حسب العدد، ويشيع الخطأ دائمًا في هذه المسألة عند كتابة العدد بالحروف، ويمكن لأي مدقق نحوي - مستعينًا بالمحلل الصرفي - أن يصحح أخطاء كتابة العدد سواء في تذكير العدد وتأنيثه أو إفراد المعدود وجمعه، فالجملة : «أصيب ما يزيد على سبع جنود» خطأ بيِّن في العربية، وصوابها: «أصيب ما يزيد على سبعة جنود»؛ لأن العدد من ٣ إلى ٩ يخالف المعدود في التذكير والتأنيث[1]. لكن ذلك يحدث دون أن يتعلم المستخدم القاعدة الهادية له في ذلك، والآلية المقترحة تزود المستخدم بتأصيل علمي وتوثيق للمعلومة اللغوية، في خطوة تهدف لتطويع التقنيات اللغوية في تعليم العربية إلكترونيًّا. ومن ثم تتعامل الآلية التعليمية مع الخطأ على النحو الآتي:

سبع جنود

اقتراح التصويب: سبعة

البحث على الإنترنت: سبعة

تحليل الخطأ: العدد سبع هنا يوافق المعدود في التذكير، والصواب أن يخالف المعدود في التذكير.

القاعدة: العدد من 3 إلى 9 يخالف المعدود في التذكير والتأنيث.

راجع لمزيد من التفصيل: العدد وتمييزه/ كتاب المصفى لـ محمد عيد [مصدر/رابط]

الشاهد الفصيح: "سَخَّرَهَا عَلَيْهِمْ سَبْعَ لَيَالٍ وَثَمَانِيَةَ أَيَّامٍ حُسُومًا" [سماع]

خطأ شائع: الخطأ في تذكير وتأنيث تمييز العدد من 3 إلى 9.

تكرار الخطأ لدى المستخدم: ثلاثة قنوات.

شكل رقم (١٩) لمعالجة المدقق النحوي التركيبَ "سبع جنود" بهدف تعليمي.

---

١ - راجع: النحو الأساسي، د. أحمد مختار، د. مصطفى النحاس، د. محمد حماسة، ص ٥٦٥ وما بعدها.

## خاتمة المبحث

سعى المبحث إلى إبراز دور تقنيات اللغة العربية الحاسوبية في التعليم الإلكتروني من خلال اتخاذ المدققات الإملائية والنحوية نموذجًا ومثالًا، ولا يقتصر الأمر على تطبيقات التدقيق فحسب بل يمتدّ إلى محركات البحث، وبرامج الترجمة الآلية، والمكانز والمعاجم أحادية اللغة (معاجم المترادفات والمتضادات) أو ثنائية اللغة (عربي - إنجليزي والعكس)، فهذه التطبيقات كذلك مما يمكن تطويعه ليفيد المستخدم العربي، ويكسبه معرفة بلغته الأم، وذلك بالتوازي مع الوظائف التقليدية لهذه الأدوات والتقنيات. وما المبحث إلا نواة وفكرة يمكن تعميمها على كثير من التطبيقات والتقنيات اللغوية.



## المبحث الثاني
## تطوير التعرف الآلي على الحروف العربية

### تمهيد

تعد عملية رقمنة مصادر المعلومات العربية أمرًا بالغ الأهمية، إذ يتم تحويل نصوص مصادر المعلومات الورقية إلى شكل إلكتروني مقروء من خلال المسح الضوئي للنصوص، وصولًا إلى نسخة إلكترونية مطابقة تماماً للنص الأصلي الموجود في صورة ( ملفات jpg – pdf على سبيل المثال).

ومع ازدياد الحاجة إلى استرجاع المعلومات والبحث في مثل هذه النصوص، زادت الرغبة في رقمنة مثل هذه الملفات، وذلك من خلال إحدى البرمجيات أو التقنيات المتخصصة في التعرف الضوئي على الحروف OCR» Optical Character Recognition“.

وتقوم هذه البرمجيات بمجموعة من العمليات والخطوات التي تستهدف التعرف على الحروف، ومن ثم تحويل الصورة إلى نص يتضمن بيانات ومعلومات يمكن استرجاعها واستخلاص ما فيها.

ومن هذه الخطوات تحسين جودة المخرجات ورفع كفاءة البرنامج عند التحويل، وذلك من خلال جود تقنية لغوية متقدمة تمكّنه من التصحيح التلقائي لأخطاء القراءة دون الاعتماد على محلل صرفي ( أو مدقق إملائي يعتمد على محلل صرفي ) وذلك بناء

على مجموعة من خصائص الحروف العربية، ومن ثم زيادة جودة هذه المخرجات، والوصول بها إلى أعلى درجة من الجودة.

ويجدر بنا قبل أن نتناول وسائل تطوير تطبيقات التعرف الآلي على الحروف العربية وعلاج مشكلاتها؛ أن نعرض للأسباب المادية التي تؤدي إلى حدوث المشكلات أثناء عملية التعرف.

## الأسباب المادية لمشكلات التعرف على الحروف العربية:

التباين الضعيف في ألوان أو «كونتراست» حروف الطباعة.

وجود البقع السوداء على النص بشكل كبير.

استخدام عدة أنواع من الخطوط ( كوفي – أندلسي – فارسي... ) في النص نفسه، أو استخدام أشكال حروف نادرة، أو حروف بخط اليد، أو حروف مكتوبة بآلة كاتبة، أو تلك الحروف التي تأخذ أشكالا رسومية «جرافيك» يكون من الصعب على برمجيات «OCR" التعرف عليها، أو أنها تخلط بينها وبين متن النص.

احتواء النص على تنسيق غاية في التعقيد مثل (وجود عدة أعمدة رأسية، ووجود إيضاحات أو هوامش أو حواشٍ في أماكن غير منتظمة، وغير ذلك)، فهذه العوامل لها تأثير سلبي على عملية التعرف الضوئي على الحروف.

احتواء النص الواحد على العديد من اللغات المختلفة.

## مشكلات التعرف على الحروف العربية، وتكمن في:

الخلط بين أشكال الحروف العربية المتشابهة (حروف الالتباس).

عدم اعتماد نموذج لغوي مدمج لحل المشكلات، والاعتماد على جهد المستخدم في التصحيح اللغوي اليدوي أو باستخدامه برامجَ التصحيح اللغوي الآلي.

وسنعمد في هذا المبحث إلى اقتراح نموذج لغوي مدمج لحل مشكلات التعرف على الحروف العربية وفق مستويي النصوص المعالجة؛ مستوى النصوص غير المضبوطة بالشكل، ومستوى النصوص المضبوطة بالشكل، وفقا لخطوات معالجة مرتبة.

## أولا : معالجة النصوص غير المضبوطة بالشكل:

وتكون من خلال اكتشاف أخطاء التعرف آليًّا ثم محاولة تصحيحها.

**طرق اكتشاف هذه المشكلات لغويًّا** ( وتكون من خلال عدة قواعد لغوية غير صرفية )

١- الحروف المتعاقبة وغير المتعاقبة.

٢ - حروف الالتباس.

٣- ترتيب الحروف العربية إحصائيًّا حسب كثرة الدوران في العربية.

٤- خصائص الحروف العربية من حيث الشكل والموضع من الكلمة.

٥- توالي الأمثال في العربية.

**وذلك على التفصيل الآتي:**

### ١- الحروف المتعاقبة وغير المتعاقبة: [١]

وهي خطوة تعتمد على الحروف العربية وإمكان اجتماعها مع بعضها، ففيها ما لا يجتمع مطلقًا بتقديم ولا تأخير في كلمة واحدة؛ كالثاء فإنها لا تجتمع بالذال والزاي والسين والصاد والضاد، وكذلك الجيم لا تجتمع بالطاء ولا بالظاء ولا بالغين ولا بالقاف ولا بالكاف، وكذلك الدال لا تجتمع بالظاء، ولا تجتمع الذال بالزاي والصاد والضاد والطاء والظاء، ولا تجتمع السين بالصاد والضاد والظاء، ولا تجتمع الصاد بالضاد ولا الظاء، ولا تجتمع الضاد بالشين والظاء، ولا تجتمع الطاء بالظاء، ولا تجتمع القاف بالغين ولا الكاف في كلمة أصلية.

والحروف الحلقية لا يجتمع بعضها ببعض، إلا الهاء فإنها تعقبها هاء الضمير وهاء التأنيث وتعقب العين أصلية كالعهد والعهر، وليس في كلمة أصلية حرفان حلقيان سوى ما تقدم من الهاء.

ولا يجتمع حرفان من هذه الخمسة وهي الهاء والطاء والعين والغين والخاء في أول كلمة سوى ما ذكر ولا في أثناء الكلمة إلا الهاء مع العين كهلع ...

١ - راجع: استخدامات الحروف العربية (معجميًّا، صوتيًّا، صرفيًّا، نحويًّا،كتابيًّا)، سليمان فياض، دار المريخ، السعودية، د.ط، ١٩٩٨. ص ١٩ وما بعدها.

ولا تتقدم الثاء الشين، وكذلك الدال لا تتقدم على زاي ولا صاد ولا طاء، وعندما عربوا كلمة «مهندز» أبدلوا الزاي سينا فقالوا «مهندس».

والذال لا تتقدم الجيم ولا السين ولا الشين ولا العين ، والشين لا تتقدمها الزاي ولا السين ولا الصاد، والطاء لا تتقدم الكاف في كلمة أصلية، والسين لا تتقدم على الدال، ويعرض الجدول الآتي للأوجه الصحيحة لتوالي الحروف العربية:

| الحروف التي لا تسبقه | حرف | الحروف التي لا تتبعه |
|---|---|---|
| ح – خ – ع – غ | أ | |
| د – ذ – ض – ط – ظ | ت | |
| ذ – ز – س– ص – ظ | ث | ذ – ز – س– ش – ص – ض – ظ |
| ص – ظ – غ – ق – ك | ج | ض – ط – غ – ق– ك |
| ث– خ – ظ – ع – غ– هـ | ح | أ – خ – ع – غ – هـ |
| ح – ظ – ع – غ – هـ | خ | أ – ح – غ – ك – هـ |
| ظ | د | ت – ذ – ز – ص– ض – ط – ظ |
| ث– د – ز – ص – ض – ط – ظ | ذ | ت – ث– د– ز– س – ش – ص – ض – ظ – غ |
| ث– د – ذ– س – ص – ط – ظ | ز | ث – ذ – س– ش – ص – ض – ظ |
| ث – ذ – ز – ص – ض – ظ | س | ث– ز – ش – ص – ض – ظ |
| ث – ذ – ز – س – ص – ض – ظ – ل | ش | ض |
| ث – د – ذ – ز – س – ض – ط – ظ | ص | ث– ج – ذ– ز– س– ش– ض – ط – ظ |
| ث – ج – د – ذ – ز – س – ش – ص – ط – ظ | ض | ت – ذ – س – ش– ص– ظ – ق |
| ج – د – ص– ظ | ط | ت– ذ– ز – ص– ض– ظ – ك |

| الحروف التي لا تسبقه | حرف | الحروف التي لا تتبعه |
|---|---|---|
| ث – ج – خ – د – ذ – ز – س – ش – ص – ض – ع – ق – ك – هـ | ظ | أ – ت – ث – د – ذ – ز – س – ص – ض – ط – ظ – غ – هـ – ي |
| ح – غ | ع | أ – ح – خ – غ |
| ج – ح – خ – ذ – ظ – ع – ق – ك – هـ | غ | ج – ح – خ – ظ – ع – ك |
| ج – ظ | ق | ج – غ – ك |
| ج – خ – ط – ظ – غ – ق | ك | |
| | هـ | ح – خ – غ |

جدول رقم (٨) للأوجه الصحيحة لتوالي الحروف العربية.

فمن خلال الجدول السابق إذن يمكن التعرف على الأوجه الصحيحة لتوالي الحروف العربية، ومن ثم يمكن حوسبة هذه القواعد وعرض مخرجات برامج التعرف الآلي عليها لأجل التحقق من صحة ودقة قراءة هذه البرمجيات بلا اعتماد على التحليل الصرفي.[١]

والسؤال المطروح هنا كيف يمكن الإفادة من الجدول السابق في اختيار أحد احتمالات الحروف المتتالية، ولذا يمكن الاعتماد على الإحصاء في ترتيب أكثر الحروف دورانًا قبل كل حرف وبعده، ويحتوي الجدول التالي على إحصاء لأكثر الحروف دورانًا مع بعضها (على مستوى الحرف الواحد)، ومن ثم يمكن التوفيق بينها وبين مخرجات البرنامج حال وجود الخطأ[٢].

١ – راجع: استخدامات الحروف العربية (معجميا، صوتيا، صرفيا، نحويا،كتابيا)، سليمان فياض، ص ١٩ وما بعدها.

٢ – راجع: استخدامات الحروف العربية (معجميا، صوتيا، صرفيا، نحويا،كتابيا)، سليمان فياض، ص ١٩ وما بعدها.

| م | أكثر الحروف التي تأتي قبله | الحرف | أكثر الحروف التي تأتي بعده | م | أكثر الحروف التي تأتي قبله | الحرف | أكثر الحروف التي تأتي بعده |
|---|---|---|---|---|---|---|---|
| | ن | أ | ب | | ر | ض | ب |
| | ر | ب | ر | | ب | ط | ر |
| | ف | ت | ر | | ع | ظ | ظ |
| | ر | ث | م | | ر | ع | ر |
| | ر– و | ج | ر | | ب | غ | م |
| | ل | ح | م | | ن | ف | ر |
| | و | خ | ر | | و | ق | ر |
| | ر | د | ر | | و | ك | ر |
| | و | ذ | ر | | ب – م | ل | م |
| | ف | ر | ب | | ر | م | ل |
| | ر | ز | ر – م | | د – ع | ن | ب |
| | ر | س | ل – م | | و | هـ | م |
| | ر | ش | ر | | ب | و | ر |
| | ص – و | ص | ل | | ل | ي | ب |

جدول رقم (٩) لأكثر الحروف دورانًا قبل كل حرف وبعده

ولكن أقصى استفادة من الجدول تتحقق بالاعتماد على جدول حروف الالتباس، وفق التفصيل الآتي:

## ٢ – حروف الالتباس:

وهي حروف متقاربة في شكلها الكتابي، وتتسبب عوامل كثيرة مادية في خلط البرنامج بينها وبين شبيهاتها من الحروف.

| الحروف الملتبسة | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| الحرف | ا | ب | ت | ج | د | ر | س | ص | ط | ع | ف | ه | ي | ــ |
| احتمال | أ | ي | ن | ح | ذ | ز | ش | ض | ظ | غ | ق | ة | ى | س |
| احتمال | إ | | | خ | | | | | | ف | ع | | ب | ش |
| احتمال | ل | | | | | | | | | ق | غ | | | ب |

جدول رقم (١٠) للحروف الملتبسة وشبيهاتها في الشكل من الحروف.

فبناءً على الخطوة الأولى الخاصة بالحروف المتعاقبة والحروف غير المتعاقبة، يفترض ألا تكون هناك كلمة فيها حرفان لا يجتمعان، وإذا ما حدث ذلك فيتم طرح البديل لأحدهما من خلال جدول إزالة الالتباس المعتمد على شكل الحرف المشبه للحرف المستبدل، واعتمادًا على خطوة لاحقة هي تقسيم الحروف العربية إحصائيًّا حسب كثرة الدوران.

وإنما اختصت هذه الخطوة بالحروف المتقاربة في الشكل فقط دون النظر إلى أي بعد لغوي؛ لاعتماد تطبيقات التعرّف على الحروف على شكل الحرف ورسمه فقط.

## ٣- ترتيب الحروف العربية إحصائيا حسب كثرة الدوران في العربية:

بعد أن أمكن -من خلال قائمة حروف الالتباس السابقة- استبدال الحروف التي تسبب مشكلة لغوية بحروف أخرى محتملة، تكون متشابهة من حيث الشكل وتسمح بها قائمة الالتباس، يمكن الاعتماد في المفاضلة بين احتمالات القائمة المتعددة على ترتيب الحروف العربية إحصائيًّا وتقسيمها حسب كثرة الدوران إلى[1]:

حروف ذائعة الاستخدام : أ - د - م - ن - ل - ف.

حروف متكررة الاستخدام : هـ - و - ب.

حروف مستخدمة : ر - ع - ف - ت - ك - د - س - ق - ح - ج.

١ - راجع: استخدامات الحروف العربية (معجميًّا، صوتيًّا، صرفيًّا، نحويًّا،كتابيًّا)، سليمان فياض، ص ١٩ وما بعدها.

حروف قليلة الاستخدام : ظ – غ – ط – ز – ث – خ – ض – ش – ص – ذ.

ويتضح من الخطوات السابقة اعتمادها على خصائص الحروف العربية الشكلية وتكراراتها في النصوص العربية، والتي تختلف عن الخطوات اللاحقة؛ إذ سيتحول التطبيق هذه المرة إلى خصائص الرسم الإملائي لبعض الحروف العربية، ومن ثم سيعتمد عليها في عملية التصحيح ورفع جودة تطبيقات التعرف الآلي على الحروف العربية.

## ٤- خصائص الحروف العربية من حيث الشكل والموضع من الكلمة (الحروف الاستهلالية - الحروف المتوسطة - الحروف المتطرفة):

للحروف العربية خصائص من حيث الشكل والموضع من الكلمة، فهناك حروف تأتي في أول الكلمة فقط، وهناك حروف تأتي في أوسطها ومنتصفها فقط، وهناك حروف متطرفة لا تأتي إلا في أواخر الكلمة، بل يتغير رسم الحرف باختلاف ترتيبه بين الحروف الأخرى بدايةً أو توسّطًا أو نهايةً. ومثال ذلك الألف أو الهمزة.

**حرف الألف أو الهمزة:**

تختلف أشكال كتابة (الألف أو الهمزة) ما بين { أ – إ – ا – ء – آ – ؤ – ئ – ى }، وتأتي هذه الصور جميعها في بداية الكلمة ماعدا { ء – ؤ – ئ – ى } ،وتأتي جميعها أيضا في وسط الكلمة ماعدا { ى }، وكذلك تأتي جميعها في نهاية الكلمة ماعدا { إ } ، وقد تجتمع أكثر من همزة في بداية الكلمة ولكن على النحو { أأ – أئ – أؤ – أ – آل } .

ومن ثم فإن أية احتمالات تخالف القواعد السابقة تعدّ خطأً؛ فعلى سبيل المثال:

١ – إذا [بدأت الكلمة] [بـ (ء) أو (ؤ) أو (ئ)] فإن [الكلمة = (خطأ)].

٢- إذا كان [وسط الكلمة] [همزة بهذا الشكل (ء)] و[كان الحرف الذي يسبقها ليس («ا» أو «د» أو «ذ» أو «ر» أو «ز» أو «و»)] ؛ فإن [الكلمة = (خطأ)].

٣- إذا كان [وسط الكلمة] [(يء)]؛ فإن [الكلمة = (خطأ)].

٤- إذا كان [في أي موضع من الكلمة] [هذا الشكل (ءء) أو (آا) أو (اآ) أو (آء) أو (آؤ) أو (آئ) أو (آأ) أو (ءآ) أو (ؤآ) أو (ئآ) أو (أآ) أو (أا) أو (اأ)] فإن [الكلمة = (خطأ)].

ومن ثم يمكن لتطبيقات التعرف على الحروف العربية الاستفادة من قواعد الرسم

الإملائي في تحسين جودتها وتصحيح مخرجاتها آليًّا، وهي الخطوة التي تقودنا نحو خطوة أخرى أو قاعدة لغوية أخرى يمكن الإفادة منها في تحسين مخرجات تطبيقات التعرف على الحروف؛ وهي قاعدة توالي الأمثال في العربية.

## ٥- توالي الأمثال في العربية:

تهرب العربية من توالي الأمثال في الحرفين أحيانا، وتهرب منه بشكل قاطع في توالي ٣ أحرف. ويمكننا حصر الحروف التي لا تتكرر مرتين، ومن ثم حذف ما سواها إذا جاء مكررًا، أما توالي ٣ حروف فممنوع في العربية منعا قاطعا مثل( للـ ) في (لالليمون) وصوابها (لليمون)، ومن ثم يمكن حذف الحرف المكرر الثالث تلقائيًّا.

## ثانيا : معالجة النصوص المضبوطة بالشكل

**تمهيد:**

من الضروري قبل الحديث عن معالجة النصوص المضبوطة بالشكل أن نمهّد لها بتمهيد؛ إذ تعد معالجة النصوص المضبوطة بالشكل الكامل مرحلة متقدمة عنها في حال النصوص غير المضبوطة بالشكل، فالنص غير المضبوط بالشكل يمثل صورة مبسطة من صور الحروف العربية التي تستطيع برامج التعرف الآلي على الحروف التعامل معها بيسر، أما النصوص المضبوطة بالشكل فهي نصوص تتسم بالتعقيد عند معالجتها بواسطة برامج التعرف الآلي على الحروف.

وتجيد بعض برامج التعرف الآلي على الحروف التعامل مع النصوص المضبوطة بالشكل، ومن هنا نشأت الحاجة إلى عمل آلية مكملة للآلية الأولى الخاصة بزيادة الجودة لبرامج التعرف الآلي على الحروف العربية، ولا تعتمد الآلية – حالها في ذلك حال الآلية الأولى – على التحليل الصرفي، وإنما على مجموعة من القواعد اللغوية الخاصة بشكل الحروف العربية، وترتيب ورودها واجتماعها في الكلام، وحركات ضبطها بالشكل.

والآلية لا تضبط النصوص بالشكل، وإنما تسعى من خلال القواعد اللغوية المختلفة إلى التأكد من تمكّن (برامج التعرف الآلي على الحروف) من التعرف على حركات التشكيل في الكتابة العربية وتصويب الخطأ حسب قواعدها، وسنعرض فيما يلي لبعض القواعد اللغوية الخاصة بالتعرف على حركات الضبط والتشكيل.

**١ – الخصائص العامة للحركات العربية (الكسرة – الضمة – الفتحة – السكون– الشدة)**

وأشكالها – على سبيل الحصر – هي: ـَ ـً ـُ ـٌ ـ ـ ـْ ـَّ ـًّ ـُّ ـٌّ ـِّ ـٍّ .

١ – الفتحة والكسرة والضمة والسكون تأتي في جميع مَواًضع الحروف من الكلمة (الحرف الأول والثاني والثالث، ... .

٢– التنوين بالفتحة والكسرة والضمة يأتي في نهايات الكلمة فقط (على الحرف الأخير) ولا يأتي التنوين في بداية الكلمة (على الحرف الأول) أو في وسطها (وقد يأتي التنوين قبل الآخر في حال التنوين بالفتحة فقط في إحدى صور كتابة التنوين بالفتحة).

٣- التنوين بالفتحة يزيد حرفا هو الألف على الكلمة غير المنتهية بتاء مربوطة أو ياء.
٤- السكون لا يأتي في الحرف الأول من الكلمة العربية، والقاعدة اللغوية في ذلك مشهورة: لا يُبتدأ بساكن .
٥- لا يتوالى ساكنان في العربية (ـــْـــْ).
٦- الشدة لا تأتي على الحرف الأول في الكلمة العربية؛ ذلك أن الشدة مكونة من حرفين: حرف ساكن+ حرف متحرك.
٧- ألف الوصل لا تضبط بالشكل.

هذه بعض الخصائص العامة للحركات العربية، لكنَّ لها خصائص أخرى إضافية – لا سميا في الشكل – إذا هي دخلت على الحروف العربية؛ كلٍّ على حدة، ولنأخذ الهمزة مثالًا تطبيقيًّا على ذلك.

**خصائص الحركات العربية مع الحروف (تطبيقا على الهمزة):**

**احتمالات تشكيل الهمزة:** [١]

يوضح الجدول التالي احتمالات ضبط الهمزة بالشكل وفق مواضعها المختلفة؛ ابتداءً وتوسطًا وانتهاءً، ويوضح كذلك اختلاف شكل الهمزة نفسها باختلاف الحركات الداخلة عليها.

| الألف والهمزة | أول الكلمة | وسط الكلمة | آخر الكلمة | ملاحظات |
|---|---|---|---|---|
| أ | تقبل من الحركات الفتحة والضمة | تقبل من الحركات الفتحة والضمة والسكون | تقبل من الحركات الفتحة والضمة والكسرة والسكون وصور التنوين كلها | |
| إ | الكسرة فقط | الكسرة فقط | لا تقع في آخر الكلمة | |

١ – راجع: الإملاء والترقيم في الكتابة العربية، عبد العليم إبراهيم، مكتبة غريب، القاهرة، ١٩٧٥م. ص ٤٥ وما بعدها.

| الألف والهمزة | أول الكلمة | وسط الكلمة | آخر الكلمة | ملاحظات |
| --- | --- | --- | --- | --- |
| ا | لا تضبط | لا تضبط | لا تضبط | |
| ء | لا تأتي في بداية الكلمة | تقبل من الحركات الفتحة والضمة والتنوين بالفتحة مع إضافة الألف بعدها | تقبل من الحركات الفتحة والضمة والكسرة وتنوين هذه الحركات جميعا | معنى ذلك أن الصور: ءِ، ءْ، ءّ لا تأتي وسط الكلمة هكذا بالضبط ذاته. وأن مجيء التنوين بالفتحة ( ـً) يأتي بعده ألف دائما. |
| ؤ | لا تأتي في بداية الكلمة | تقبل من الحركات الفتحة وا لضمة والسكون والتنوين بالفتحة مع إضافة الألف بعدها | تقبل من الحركات الفتحة و الضمة والكسرة وتنوين الضم والكسر | يلاحظ عدم تشديد الهمزة المتوسطة في أي من حالتها |
| ئ | لا تأتي في بداية الكلمة | تقبل من الحركات الفتحة وا لضمة والسكون والتنوين بالفتحة مع إضافة الألف بعدها | تقبل من الحركات الفتحة و الضمة والكسرة والسكون وتنوين الضم والكسر | |
| ى (ألف ممدودة) | لا تأتي في بداية الكلمة | لا تأتي في وسط الكلمة | لا تضبط | قبلها مفتوح دائما وصوره: ـَى – ـًى – ـَّى – ـًّى |

جدول رقم (١١) احتمالات ضبط الهمزة بالشكل وفق مواضعها المختلفة

**ومن الممكن زيادة في الإيضاح؛ تفصيل الجدول السابق في عدة جداول أخرى على النحو الآتي:**

## احتمالات تشكيل الهمزة في أول الكلمة:

| شكلها | تشكيلها |
|---|---|
| أ | إما أن تكون مضمومة أو مفتوحة |
| إ | لابد أن تكون مكسورة |

جدول رقم (١٢) لاحتمالات تشكيل الهمزة في أول الكلمة

ولا تأتي همزة القطع في بداية الكلمة بهذا الشكل ( ؤ ) ولا بهذا (ئ) ولا بهذا (ء)

**الهمزة على ألف (أ):**

| شكلها | تشكيلها | تشكيل الحرف الذي قبلها | الحرف الذي يسبقها |
|---|---|---|---|
| أ | فتحة | سكون | |
| أ | فتحة أو سكون | فتحة | |
| أ | فتحة أو ضمة | كسرة | لام الجر أو باء الجر |

جدول رقم (١٣) لاحتمالات تشكيل الهمزة

**الهمزة تحت الألف (إ):**

| شكلها | تشكيلها | تشكيل الحرف الذي يسبقها | الحرف الذي يسبقها |
|---|---|---|---|
| إ | كسرة | فتحة | لام التوكيد أو واو العطف أو فاء العطف أو الكاف الجارة أو همزة الاستفهام |
| إ | كسرة | كسرة | اللام الجارة أو الباء الجارة |

جدول رقم (١٤) لاحتمالات تشكيل الهمزة المكسورة

## احتمالات تشكيل الهمزة في وسط الكلمة[1]:

### الهمزة التي على نبرة (ئـ):

| شكلها | تشكيلها | تشكيل الحرف الذي قبلها | الحرف الذي قبل الهمزة | الحرف الذي بعد الهمزة |
|---|---|---|---|---|
| ئـ | كسرة | كسرة أو ضمة أو فتحة أو سكون أو | | |
| ئـ | كسرة | ألف ممدودة ما قبلها مفتوح | | |
| ئـ | كسرة أو ضمة أو فتحة أو سكون | كسرة | | |
| ئـ | فتحة | سكون | ياء | مثال : هيئة |
| ئـ | تنوين مفتوح | سكون | | ألف التنوين (دفئا) |
| ئـ | فتحة | كسرة | | ألف الاثنين (دفِئَا) |
| ئـ | فتحة | فتحة | ب – ت – ث – ج – ح – خ – س – ش – ص – ض – ط – ظ – ع – غ – ف – ق – ك – ل – م – ن – هـ | ألف مد مكافئات |
| ئـ | ضمة | بلا حركة (ياء مدية) | ياء ممدودة | واو ممدودة بريئون |

١ – راجع: الإملاء والترقيم في الكتابة العربية، عبد العليم إبراهيم، ص ٥٤، ٥٥.

| شكلها | تشكيلها | تشكيل الحرف الذي قبلها | الحرف الذي قبل الهمزة | الحرف الذي بعد الهمزة |
|---|---|---|---|---|
| ئـ | ضمة | ضمة أو سكون | ب – ت – ث – ج – ح – خ – س – ش – ص – ض – ط – ظ – ع – غ – ف – ق – ك – ل – م – ن – هـ | واو ممدودة مسئول / شئون |
| ئـ | سكون | كسرة | ألف وصل في أول الكلمة | ائذن |

جدول رقم (١٥) لاحتمالات تشكيل الهمزة التي على نبرة

**الهمزة التي على واو (ؤ):**

| شكلها | تشكيلها | تشكيل الحرف الذي قبلها | الحرف الذي يسبق الهمزة | الحرف الذي بعد الهمزة |
|---|---|---|---|---|
| ؤ | ضمة | فتحة أو سكون | | |
| ؤ | ضمة | ألف ممدودة ما قبلها مفتوح | | |
| ؤ | ضمة أو فتحة أو سكون | ضمة | | |
| ؤ | سكون | ضمة | ألف وصل في أول الكلمة | اؤمر |

جدول رقم (١٦) لاحتمالات تشكيل الهمزة التي على واو

### الهمزة على السطر (ء):

| شكلها | تشكيلها | تشكيل الحرف الذي قبلها | الحرف الذي يسبقها | الحرف الذي يلي الهمزة |
|---|---|---|---|---|
| ء | فتحة | | ألف ممدودة | ألف تنوين (آخر حرف) |
| ء | فتحة | | ألف ممدودة | |
| ء | تنوين | | د- ذ – ر- ز – و | ألف تنوين (آخر حرف) |
| ء | فتحة | | د- ذ – ر- ز – و | ألف تنوين (آخر حرف) |
| ء | فتحة | ساكن | د- ذ – ر- ز – و | ألف ممدودة |
| ء | فتحة | | واو ممدودة | |
| ء | ضمة | ضمة | واو ممدودة | |
| ء | ضمة | | د- ذ – ر- ز – و | واو ممدودة |

جدول رقم (١٧) لاحتمالات تشكيل الهمزة التي على على السطر

### احتمالات تشكيل الهمزة في آخر الكلمة:

| شكلها | تشكيل الحرف الذي يسبقها | الحرف الذي يسبقها |
|---|---|---|
| ئ | كسرة | |
| ؤ | ضمة | |
| أ | فتحة | |
| ء | سكون | |
| ء | | حرف مد (و – ا –ي ) |

جدول رقم (١٨) لاحتمالات تشكيل احتمالات تشكيل الهمزة في آخر الكلمة

## الخاتمة

عرضنا في هذا المبحث لبعض وسائل تطوير التعرف الآلي على الحروف العربية، من خلال أداة لغوية غير صرفية، تتضمن قواعد الاكتشاف المبكر للأخطاء الناتجة عن التعرف الآلي على الكلام باللغة العربية دون الاعتماد على محلل صرفي، وذلك على مستوى النصوص المضبوطة بالشكل والنصوص غير المضبوطة بالشكل، وهي عملية ستسهم في تحسين جودة المخرجات ورفع كفاءة البرنامج، إذ توفر تلك الآلية المقترحة للمستخدم تقنية لغوية عربية متقدمة تمكنه من التحقق والتصحيح التلقائي لأخطاء القراءة دون الاعتماد على محلل صرفي في البداية، حيث يمكن - بالطبع في خطوة تالية - استخدام المحلل الصرفي لمعالجة مخرجات آليتنا؛ توفيرًا للوقت والجهد في الترجيح بين الاحتمالات المتعددة (اللبس الصرفي)، فالكثير من احتمالات الخطأ قد حُذفت، هذا إن لم يكن اللبس قد زال قبل مرحلة التحليل الصرفي.



# الفصل الثالث
# آفاق استخدام بعض التقنيات اللغوية

## المبحث الأول
## مراجعة النص القرآني وتدقيقه آليًّا

### تمهيد:

مع التقدم الكبير في استخدام وسائل التواصل الاجتماعي واستخدام الهواتف النقالة الذكية، شاعت النُّسخ الإلكترونية المختلفة للقرآن الكريم –ومنها على سبيل المثال النسخ التي تعرضها متاجر (Google Play ، Apple Store) – وشاع الخطأ فيها حتى غدت الحاجة ملحّةً إلى آلية لغوية تمكننا من مراجعة النص القرآني كلمةً كلمةً بل حرفًا حرفًا وحركةً حركةً، والتصديق على صحته، وأتت فكرة هذا المبحث لتقترح آلية لغوية هدفها التدقيق الإملائي واللغوي للنص القرآني في نُسخه الرقمية برسمه العثماني.

وتعتمد الآلية المقترحة في هذا المبحث في عملها على عدد من الأدوات الحاسوبية المقترحة التي اقترحت في مبحث سابق لتطوير آليات التعرف على الحروف:

### ١ – الأدوات الإحصائية من خلال تطبيقات الإحصاء والفهرسة اللغوية:

على مستوى الحرف.

على مستوى الكلمة.

على مستوى نهايات الآي وصفحات المصحف.

### ٢- الأدوات اللغوية من خلال عدد من الخصائص والقواعد اللغوية الحاسوبية:

خصائص الحروف العربية.

خصائص حركات الضبط بالشكل (حركات التشكيل).

خصائص الحركات العربية مع بعض الحروف (تطبيقًا على الهمزة).

### ٣ - أدوات معالجة الرسم العثماني آليًّا من خلال حوسبة قواعد الرسم العثماني وأوجه الاختلاف بينه وبين الرسم الكتابي للعربية:

- حذف بعض الحروف.
- زيادة بعض الحروف.
- كتابة الهمزة على غير القياس.
- استبدال حرف بحرف.
- الفصل والوصل بين أجزاء كلمات القرآن.

ويمكن أن تصل كفاءة هذه الآلية إلى ما يقرب من ١٠٠ ٪ ، نظرًا لتطبيقها وتطبيق قواعدها اللغوية والإحصائية على القرآن الكريم فقط بوصفه نصًّا مغلقًا (Closed Text)، غير قابل للزيادة أو النقص أو التغيير أو التبديل، وهو ما يعكس سهولة حوسبته ومعالجته آليًّا ودقة النتائج وصدقها في الوقت نفسه.

## المقدمة

يهدف هذا المبحث إلى التعامل مع النص القرآني الرقمي بعيدًا عن النص المكتوب (أو المرسوم) بخط الخطاط في المصاحف الورقية، فلذلك طرق أخرى لمعالجته من برامج التعرف الضوئي على الحروف (OCR)[1]، كما يهدف إلى رفع جودة المراجعة لنص القرآن العظيم، والتأكيد على دقة النسخ الإلكترونية والمطبوعة لهذا الكتاب الذي لا يأتيه الباطل من بين يديه ولا من خلفه. والمبحث الذي بين أيدينا وإن قصد إلى آلية للتعامل مع الرسم العثماني للقرآن الكريم فإنه لا يهمل الرسم الإملائي الحديث لكلمات القرآن ويربطها مباشرة بمقابلاتها في الرسم العثماني. أما الأدوات التي تعتمدها الآلية في عملها فهي –على التفصيل– كما يأتي:

---

١ – راجع: تطوير التعرف الآلي على الحروف العربية، عمرو جمعة عبد الرسول، الندوة الدولية السادسة لعلوم وهندسة الحاسوب، ٢٠١٠ تونس. ص ٣٦٥ ، وراجع أيضا: المعايير اللغوية والفنية لتقييم برامج التعرف الضوئي على الحروف العربية، عمرو جمعة عبد الرسول، مؤتمر جمعية هندسة اللغة العاشر، ٢٠١٠ القاهرة، ص ١١٠.

## الأدوات الحاسوبية المقترحة التي تتضمنها آلية تدقيق المصحف:

## الأدوات الإحصائية:

**مقدمة في التعريف بتطبيقات الإحصاء والفهرسة اللغوية[1]:**

تهتمّ تطبيقات الفهرسة اللغوية بعملية التحليل الإحصائي لمفردات النصوص المعالجة من خلال إعداد فهارس هجائية لمفردات هذه النصوص، وتحديد عدد مرات تكرارها، وتُعرض هذه الفهارس في صورة قائمة مرتبة وفق الترتيب الهجائي التصاعدي أو التنازلي، أو وفق عدد تكرار الكلمات تصاعديًّا أو تنازليًّا، أو وفقًا لطول الكلمة تصاعديًّا أو تنازليًّا.

ومن أمثلة هذه التطبيقات[2] Concordance Version ٢ , ٣ [3] ، و Textanz [4].

**برنامج Concordance الإحصائي:**

يدعم البرنامج اللغة العربية وحروفها، وتُدرج ملفات النصوص فيه بامتداد(.txt) فقط، إذ يقوم البرنامج بفهرسة الكلمات الموجودة بهذه النصوص. والكلمة عنده: مجموعة من الحروف المتصلِ بعضُها ببعض، الواقعةِ بين مسافتين قبلها وبعدها، ويعرض البرنامج النص في شكل عمود واحد من الكلمات Headword ويوضع أمام كل كلمة عدد تكرار هذه الكلمة في النص NO. ، ثم يعرض نسبة مئوية لهذا التكرار بالمقارنة مع النص كله.

برنامج Textanz الإحصائي:

يتميز برنامج Textanz عن برنامج Concordance بمعالجة التراكيب – إضافة إلى معالجة الكلمات – فباختيار وظيفة Phrase ، يعرض البرنامج التراكيب المتكررة في النص مصحوبة بعدد مرات تكرارها ، وعدد كلمات التركيب.

ويمكن للبرنامج تظليل الكلمات محل البحث في مواضعها في النص، ويتميز

---

١ – راجع: مدونة معجم عربي معاصر، معالجة لغوية حاسوبية، المعتز بالله السعيد طه، رسالة ماجستير، كلية دار العلوم، جامعة القاهرة، ١٤٢٨هـ/ ٢٠٠٧م، ص٦٩ وما بعدها. وراجع أيضا: اللغة العربية المعاصرة في المجال السياسي، دراسة لغوية حاسوبية من واقع ذخيرة لغوية، هبة سيد راشد، ص ٣٧، ٣٨ بتصرف.

٢ – سيعرض الكتاب للتعريف بالتطبيقين في الفصل الثالث لمناسبة المقام هناك للاستفاضة في التعريف بهما.

٣ – http: //www.concordancesoftware.co.uk/

٤ – http: //www.cro-code.com /textanz.jsp/

البرنامج بميزة أخرى وهي Words forms وتختص بعرض الصور المختلفة للكلمة الواحدة، حيث يعرض للكلمات متشابهة الحروف في النص الواحد، فكلمات مثل: ﴿إِلَٰهٌ﴾، ﴿وَإِلَٰهُنَا﴾، ﴿وَإِلَٰهُكُمْ﴾ تأتي تحت ساق الكلمة «إله»، ويمكن التحكم في عدد حروف التشابه بتحديد أقصى عدد لحروف الكلمات المتشابهة أو أقل عدد لحروف الكلمات المتشابهة.

**كيفية استخدام برامج الفهرسة والإحصاء في اكتشاف الأخطاء على مستوى الحروف والكلمات والآيات والصفحات:**

تتلخص فكرة استخدام تطبيقات الفهرسة الآلية في تدقيق النص القرآني في الاعتماد على فهرسة نسخة إلكترونية أصلية ومنقحة ومراجَعةٍ أكثر من مرة من القرآن الكريم برسمه العثماني، بوصفه نسخة أصلية يمكن القياس عليها وعدّها أصلاً، ثم مقارنة أية نسخ جديدة بها، وبإحصاءاتها المستخرجة من تطبيقات الفهرسة الآلية، ومن ثم فإن خطوات عملنا هي:

توفير نسخة إلكترونية أصلية ومنقحة ومراجعة أكثر من مرة من القرآن الكريم في رسمه العثماني، وبامتداد .txt

عمل عدد من الإحصاءات على هذه النسخة الأصلية (سيأتي شرحها بالتفصيل).

عمل هذه الإحصاءات نفسها على النسخة المراد مراجعتها، ذات الامتداد .txt

مطابقة نتائج إحصاءات النسخة المراد مراجعتها مع نتائج إحصاءات النسخة الأصلية.

اكتشاف الاختلافات بين النسختين والتصحيح وفقًا للنسخة الأصلية القديمة.

الإحصاءات التي يمكن استخراجها باستخدام برامج الفهرسة الآلية والتي تفيد في مراجعة النص وتدقيقه آليًّا:

أ- إحصاء بالمجموع الإجمالي لعدد كلمات القرآن الكريم كاملًا في رسمه العثماني.

ب- إحصاء بالمجموع الإجمالي لعدد كلمات كل سورة من سور القرآن الكريم على حدة في رسمه العثماني.

ج- إحصاء بالمجموع الإجمالي لتردّد كلمات القرآن الكريم كاملًا في رسمه العثماني.

د- إحصاء بالمجموع الإجمالي لتردّد كلمات كل سورة من سور القرآن الكريم على حدة في رسمه العثماني.

هـ- إحصاء بالمجموع الإجمالي لتراكيب القرآن الكريم كاملًا في رسمه العثماني.

و- إحصاء بالمجموع الإجمالي لتردّد تراكيب كل سورة من سور القرآن الكريم على حدة في رسمه العثماني.

## مستويات التدقيق باستخدام تطبيقات الإحصاء والفهرسة اللغوية

### ١- التدقيق على مستوى حروف الكلمات القرآنية:

ولا تقتصر عملية المصادقة والتدقيق على مستوى الكلمات القرآنية، بل تتجاوزها إلى حروف هذه الكلمات من خلال العديد من الإحصاءات والفهارس لحروف الكتاب العزيز، على النحو الآتي:

أ- الحروف غير المتجانسة في العربية (الحروف عديمة الائتلاف)

في العربية حروف لا تجتمع مطلقًا في كلمة واحدة، لا بتقديم ولا تأخير، وقد سبق عرض ذلك في مبحث تطوير آليات التعرف على الحروف العربية، وهي كما يأتي في الجدول المبيّن[1].

| الثنائيات - عديمة الائتلاف | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| س ث | س ذ | س ص | س ض | س ظ | س ش | ص ج | ص ش | ط ز | د ص | ض ق | د ض |
| ث س | ذ س | ص س | ض س | ظ س | ث س | ض ط | ط ض | ض ظ | ظ ض | ض ش | ش ض |
| ث ذ | ث ز | ث ص | ث ض | ث ظ | ذ س | ظ ط | ط ظ | ظ ج | ج ظ | ظ د | د ظ |
| ذ ث | ز ث | ص ث | ض ث | ظ ث | س ث | ظ ح | ظ ق | ظ ش | ظ خ | خ غ | غ خ |
| ث ش | ذ ش | ذ غ | ز ش | ز ض | س ذ | ج ط | ج غ | ج ق | ع خ | د ز | د ط |
| ذ ز | ذ ص | ذ ض | ذ ط | ذ ظ | س ز | ط ج | غ ج | ق ج | ع غ | غ ع | ق غ |
| ز ذ | ص ذ | ض ذ | ط ذ | ظ ذ | ز س | ح خ | خ ح | ح ع | ع ح | ح غ | غ ح |
| ص ض | ض ص | ص ط | ط ص | ص ظ | ظ ص | ز ص | ص ز | ز ط | ظ ز | | |

جدول رقم (٢٠) لثنائيات الحروف غير المتجانسة في العربية

١ http://www.diwa-alarab.com/IMG/pdf/Is_hamaatUolamaaAltaumieat1-1.pdf في 2014-1-12.

فهذه الثنائيات من الحروف العربية – كما ذكرنا من قبل – تتنافر فيما بينها ولا تجتمع في كلمة عربية أبدًا، ومن ثم فإن اجتماع مثل هذه الحروف غير المتجانسة في كلمة واحدة –بله في القرآن الكريم– يدل على وقوع الخطأ في الكلمة.

ويوضح الجدول الآتي بطريقة أخرى الحروف التي تتآلف والأخرى التي تتنافر، من خلال حصر الحروف التي لا تسبق كل حرف من الحروف الهجائية العربية، وكذا الحروف التي لا تلي كل حرف من حروف الأبجدية العربية [1].

| الحروف التي لا تسبق | حرف | الحروف التي لا تلي |
|---|---|---|
| ح – خ – ع – غ | أ | |
| د – ذ – ض – ط – ظ | ت | |
| ذ – ز – س – ص – ظ | ث | ذ – ز – س – ش – ص – ض – ظ |
| ص – ظ – غ – ق – ك | ج | ض – ط – غ – ق – ك |
| ث – خ – ظ – ع – غ – هـ | ح | أ – خ – ع – غ – هـ |
| ح – ظ – ع – غ – هـ | خ | أ – ح – غ – ك – هـ |
| ظ | د | ت – ذ – ز – ص – ض – ط – ظ |
| ث – د – ز – ص – ض – ط – ظ | ذ | ت – ث – د – ز – س – ش – ص – ض – ظ – غ |
| ث – د – ذ – س – ص – ط – ظ | ز | ث – ذ – س – ش – ص – ض – ظ |
| ث – ذ – ز – ص – ض – ظ | س | ث – ز – ش – ص – ض – ظ |
| ث – ذ – ز – س – ص – ض – ظ – ل | ش | ض |
| ث – د – ذ – ز – س – ض – ط – ظ | ص | ث – ج – ذ – ز – س – ش – ض – ط – ظ |
| ث – ج – د – ذ – ز – س – ش – ص – ط – ظ | ض | ت – ذ – س – ش – ص – ظ – ق |
| ج – د – ص – ظ | ط | ت – ذ – ز – ص – ض – ظ – ك |

---

١ – راجع: استخدامات الحروف العربية (معجميًّا، صوتيًّا، صرفيًّا، نحويًّا،كتابيًّا)، سليمان فياض، ص ١٩ وما بعدها.

| الحروف التي لا تسبق | حرف | الحروف التي لا تلي |
|---|---|---|
| ث – ج – خ – د – ذ – ز – س – ش – ص – ض – ع – ق – ك – هـ | ظ | أ – ت – ث – د – ذ – ز – س – ص – ض – ط – ظ – غ – هـ – ي |
| ح – غ | ع | أ – ح – خ – غ |
| ج – ح – خ – ذ – ظ – ع – ق – ك – هـ | غ | ج – ح – خ – ظ – ع – ك |
| ج – ظ | ق | ج – غ – ك |
| ج – خ – ط – ظ – غ – ق | ك | |
| | هـ | ح – خ – غ |

جدول رقم (٢١) لحالات اجتماع الحروف العربية مع بعضها

ويُفاد من الجدولين السابقين التعرّفُ على الأوجه الصحيحة لتوالي الحروف العربية، ومن ثم حوسبة هذه القواعد من خلال آلية مراجعة وتدقيق حاسوبية، تعتمد على الإحصاء والفهرسة.

**ب- عدد حروف القرآن الكريم وأعداد تكراراتها:**

للإحصاء دور مهم في تدقيق حروف القرآن الكريم وتدقيق تكراراتها فيه، فمجمل حروف القرآن ٣٣٠٧٣٣ ، وأية نتيجة أخرى تنتج عن تطبيقات الإحصاء والفهرسة تعني وجود خطأ ما. ثم إن من الممكن التعامل مع عدد حروف سور القرآن سورة سورة، إمعانًا في المبالغة في تدقيق النص القرآني.

وفي مرحلة أخرى متقدمة يمكن التصديق على عدد تكرار حروف العربية في القرآن في النص القرآني كله، وتأكيد صحتها في النسخة المراد مراجعتها، من خلال مقارنة أعداد تكراراتها في النسخة المراد مراجعتها بأعداد تكراراتها في النسخة الأصلية الموثقة. ويبين الجدول الآتي حروف القرآن الكريم مرتبة من الأكثر تكرارًا إلى الأقل تكرارًا.

| Percentage | Frequency | Letter | .No |
|---|---|---|---|
| ١٣,٠٩٩ | ٤٣٣٢٤ | ا | ١ |
| ١١,٥٤٧ | ٣٨١٩١ | ل | ٢ |
| ٨,٢٤٦ | ٢٧٢٧١ | ن | ٣ |
| ٨,٠٨٤ | ٢٦٧٣٥ | م | ٤ |
| ٧,٥٠٢ | ٢٤٨١٣ | و | ٥ |
| ٦,٦٤٥ | ٢١٩٧٧ | ي | ٦ |
| ٤,٤٩٠ | ١٤٨٥٠ | ه | ٧ |
| ٣,٧٥٠ | ١٢٤٠٣ | ر | ٨ |
| ٣,٤٧٤ | ١١٤٩١ | ب | ٩ |
| ٣,١٧٥ | ١٠٥٠١ | ت | ١٠ |
| ٣,١٧٤ | ١٠٤٩٧ | ك | ١١ |
| ٢,٨٤٤ | ٩٤٠٥ | ع | ١٢ |
| ٢,٧٥٧ | ٩١١٩ | أ | ١٣ |
| ٢,٦٤٥ | ٨٧٤٨ | ف | ١٤ |
| ٢,١٢٧ | ٧٠٣٤ | ق | ١٥ |
| ١,٨١٨ | ٦٠١٢ | س | ١٦ |
| ١,٨١١ | ٥٩٩١ | د | ١٧ |
| ١,٥٣٨ | ٥٠٨٦ | إ | ١٨ |
| ١,٤٩١ | ٤٩٣٢ | ذ | ١٩ |
| ١,٢٥٢ | ٤١٤٠ | ح | ٢٠ |
| ١,٠٠٣ | ٣٣١٧ | ج | ٢١ |
| ٠,٧٨٧ | ٢٦٠٣ | ى | ٢٢ |

| .No | Letter | Frequency | Percentage |
|---|---|---|---|
| ٢٣ | خ | ٢٤٩٧ | ٠,٧٥٥ |
| ٢٤ | ة | ٢٣٦٣ | ٠,٧١٤ |
| ٢٥ | ش | ٢١٢٤ | ٠,٦٤٢ |
| ٢٦ | ص | ٢٠٧٢ | ٠,٦٢٦ |
| ٢٧ | آ | ١٧٣٠ | ٠,٥٢٣ |
| ٢٨ | ض | ١٦٨٦ | ٠,٥١٠ |
| ٢٩ | ز | ١٥٩٩ | ٠,٤٨٣ |
| ٣٠ | ء | ١٥٢٥ | ٠,٤٦١ |
| ٣١ | ث | ١٤١٤ | ٠,٤٢٨ |
| ٣٢ | ط | ١٢٧٣ | ٠,٣٨٥ |
| ٣٣ | غ | ١٢٢١ | ٠,٣٦٩ |
| ٣٤ | ئ | ١١٥١ | ٠,٣٤٨ |
| ٣٥ | ظ | ٨٥٣ | ٠,٢٥٨ |
| ٣٦ | ؤ | ٧٨٥ | ٠,٢٣٧ |

جدول رقم (٢٢) لحروف القرآن الكريم مرتبة من الأكثر تكرارًا إلى الأقل تكرارًا

بعد التحقق من تكرارات الحروف في القرآن الكريم كاملًا والتصديق على صحتها في النسخة المراد مراجعتها يمكن الانتقال إلى مرحلة أخرى من التصديق على تكرارات الحروف العربية، لكن على مستوى كل سورة على حدة، ومن ثم ضمان تأكيد صحة عدد تكرارات الحروف في كل سورة بعد ضمان التأكد من التكرارات في النص القرآني كله.

### ج- توالي الأمثال (الحروف المتماثلة) في القرآن الكريم:

ترفض العربية بشكل قاطع توالي الأمثال إذا توالت ثلاثة أحرف، وتكرهه في العربية

إذا توالى حرفان، فالكتابة الإملائية الصحيحة لكلمة (لالليمون) [ل+ال+ليمون] - على سبيل المثال - تكون هكذا : (لليمون)، درءًا لتوالي الأمثال [ل(لام الجر) +ل(لام التعريف) +ل(لام الكلمة)] [1]. فليس ثمة في العربية ثلاثة أحرف متماثلة في كلمة واحدة، والتأكد من خلو نسخة القرآن الكريم المراد مراجعتها من توالي ثلاثة أمثال في كلمة واحدة هي أولى خطوات هذه المرحلة.

أما الحروف التي تتماثل في القرآن الكريم على مستوى حرفين فيمكن إحصاء عدد مرات تكراراتها مع أمثالها في القرآن الكريم كاملًا، وفي كل سورة من سوره، وهي خطوة أخرى من خطوات هذه المرحلة. والجدول الآتي يوضح أعداد تكرارات تعاقب الحروف العربية في القرآن الكريم مرتين مباشرة بلا فاصل في كلمة واحدة كما في (للـ، مم، نن ، دد، جج، ... ):

| حرف اللام ٣٢١٥ مرة. | حرف الصاد ١٠١ مرة. | حرف الدال ٣٢ مرة. | حرف الكاف ٤ مرات. |
| --- | --- | --- | --- |
| حرف الميم ٢١٨ مرة. | حرف الفاء ٦٥ مرة. | حرف الياء ٢٦ مرة. | حرف الطاء ٣ مرات. |
| حرف الألف ١٩٠ مرة. | حرف الواو ٤٩ مرة. | حرف الراء ١٨ مرة. | حرف الجيم مرتين. |
| حرف التاء ١٤٩ مرة. | حرف الهاء ٤٨ مرة. | حرف السين ١٤ مرة. | حرف الزاي مرتين. |
| حرف النون ١٤٢مرة. | حرف الباء ٤٥ مرة. | حرف القاف ٥ مرات. | حرف الضاد مرتين. |

جدول رقم (٢٣) لعدد تكرارات توالي الأمثال في القرآن الكريم

الخطوة التالية والأخيرة تتضمن حصر الحروف التي لا تتكرر مرتين في القرآن الكريم، ومن ثم تخطئتها إذا جاءت مكررة في النسخة المراجَعة، وهي على وجه الحصر: حروف الثاء والحاء والخاء والذال والشين والظاء والعين والغين والهمزة، فلم تقع جميعها متتابعة في القرآن الكريم كله[2].

١ - راجع: الإملاء والترقيم في الكتابة العربية، عبد العليم إبراهيم، ص ٧٨ .

٢ - راجع: التقاء الحروف المتماثلة في القرآن الكريم، د.محمد زكي خضر ، مجلة الفرقان، العدد ٥٠، آذار ٢٠٠٦. على

## ٢- التدقيق على مستوى الكلمات من خلال تحويل النص القرآني إلى قائمة كلمات Text to list :

تهدف هذه الخطوة إلى تحويل كلمات النسخة الأصلية المراجَعة من النص القرآني إلى قائمة كلمات طويلة، أو إلى جدول من عمود رأسي واحد، ثم تسهيل عملية المقارنة الآلية بينه وبين النسخة المراد مراجعتها، والتي ستُحوَّل أيضًا إلى قائمة عمودية من الكلمات، أو إلى جدول ذي عمود واحد، ويمكن حينئذ تحديد الخلايا الأفقية غير المتساوية في الخليتين الأفقيتين المتقابلتين في كل عمود منهما.

والشكل الآتي يوضح جانبًا من هذه المقارنة بين نسختين إحداهما بالرسم العثماني والأخرى رقمية يراد تصحيحها.

| ذَلِكَ | ذَٰلِكَ |
|---|---|
| الْكِتَابُ | ٱلْكِتَٰبُ |
| لا | لَا |
| رَيْبَ | رَيْبَ |
| فِيهِ | فِيهِ |
| هُدًى | هُدًى |
| لِلْمُتَّقِينَ | لِّلْمُتَّقِينَ |
| (٢) | ٢ |

جدول رقم (٢٤) للمقارنة بين نسختين في عمودين، إحداهما مراجَعة وصحيحة، والأخرى يراد تصحيحها.

الرابط http: / /www.al-mishkat.com /khedher /wp-admin /Papers /paper18.pdf في 2014-1-12.

### ٣ – التدقيق على مستوى رؤوس الآي ونهايات الصفحات في مصحف المدينة من خلال تحويل النص القرآني إلى قائمة كلمات Text to list :

بعد تحويل كلمات النسخة الأصلية المراجعة من النص القرآني إلى قائمة كلمات طويلة أو إلى جدول من عمود رأسي واحد وإجراء عملية المقارنة الآلية بينه وبين النسخة المراد مراجعتها، وهي التي حُوِّلت إلى قائمة عمودية من الكلمات أو إلى جدول ذي عمود واحد؛ يمكن بعد ذلك تحديد الخلايا الأفقية التي تحمل أرقام رؤوس الآيات غير المتساوية في الخليتين الأفقيتين المتقابلتين في كل عمود منهما.

| بِسْمِ | بِسْـــمِ | الْحَمْدُ | ٱلْحَـمْدُ | | |
|---|---|---|---|---|---|
| اللَّهِ | ٱللَّهِ | لِلَّهِ | لِلَّهِ | | |
| الرَّحْمَنِ | ٱلرَّحْمَٰنِ | رَبِّ | رَبِّ | الرَّحْمَنِ | ٱلرَّحْمَٰنِ |
| الرَّحِيمِ | ٱلرَّحِيمِ | الْعَالَمِينَ | ٱلْعَٰلَمِينَ | الرَّحِيمِ | ٱلرَّحِيمِ |
| (١) | ١ | (٢) | ٢ | (٣) | ٣ |
| | | إِيَّاكَ | إِيَّاكَ | | |
| مَالِكِ | مَٰلِكِ | نَعْبُدُ | نَعْبُدُ | اهْدِنَا | ٱهْدِنَا |
| يَوْمِ | يَوْمِ | وَإِيَّاكَ | وَإِيَّاكَ | الصِّرَاطَ | ٱلصِّرَٰطَ |
| الدِّينِ | ٱلدِّينِ | نَسْتَعِينُ | نَسْتَعِينُ | الْمُسْتَقِيمَ | ٱلْمُسْتَقِيمَ |
| (٤) | ٤ | (٥) | ٥ | (٦) | ٦ |

جدول رقم (٢٥) للمقارنة بين رؤوس الآيات في سورة الفاتحة في النسختين؛ الأصلية والمراد مراجعتها وتدقيقها

أما تدقيق نهايات الصفحات في مصحف المدينة، فيمكن إجراؤه من خلال عملية (ترميز) وتعني وضع علامة معرّفة في بدايات ونهايات كل صفحة من صفحات مصحف المدينة المنورة، وتحديد أول كلمتين في كل صفحة من صفحات القرآن الكريم في مصحف المدينة وآخر كلمتين بها، وربط ذلك كله برقم كل صفحة، وبناء قاعدة بيانات بذلك لعقد مقارنة بين النسخة المراد مراجعتها وقاعدة البيانات هذه.



شكل رقم (٢٠) لصفحتين متقابلتين من مصحف المدينة المنورة لشرح العلاقة بين رقم الصفحة وكلمات البداية فيها وكلمات نهاية الصفحة

فالصفحة ٣ من مصحف المدينة المنورة تبدأ ﴿إِنَّ ٱلَّذِينَ كَفَرُواْ﴾ (البقرة: ٦)، وتنتهي بـ ﴿وَمَا كَانُواْ مُهۡتَدِينَ﴾ (البقرة: ١٦)، وكذلك الصفحة ٤ من مصحف المدينة تبدأ بـ﴿مَثَلُهُمۡ كَمَثَلِ ٱلَّذِي﴾ (البقرة: ١٧) وتنتهي بـ﴿أُعِدَّتۡ لِلۡكَٰفِرِينَ﴾ (البقرة: ٢٤)، ومن ثم تعدّ هذه النسخة المرمَّزة أصلًا يمكن المقارنة به لاكتشاف الأخطاء في بدايات كل صفحات القرآن ونهاياتها حسب مصحف المدينة المنورة.

## ٢- الأدوات اللغوية:

وتتضمن كيفية الإفادة من الدراسات اللغوية الحاسوبية في عملية التدقيق لنص القرآن الكريم، من خلال بعض القواعد اللغوية الحاسوبية، كالقواعد النحوية والصرفية وقواعد الإعراب والضبط بالشكل، والاعتماد عليها بوصفها وسيلة من وسائل التدقيق والمصادقة لنص القرآن الكريم، وقد ذكرنا بعضها عند الحديث عن تطوير آليات التعرف على الحروف العربية.

**أ – خصائص الحروف العربية:**

لبعض الحروف العربية خصائص من حيث الشكل والموضع في الكلمة، فهناك حروف يختلف رسمها باختلاف مكانها، فترسم في أول الكلمة فقط برسم يختلف عن رسمها في وسط الكلمة عنه في آخرها، فرسم الحرف يتغير باختلاف ترتيبه بين الحروف الأخرى استهلالًا أو توسطًا أو تطرفًا. وقد ضربنا مثالًا بالهمزة عند الحديث عن تطوير تقنيات التعرف على الحروف العربية، فأشكال كتابة (الهمزة) في العربية تختلف حسب موقعها في الكلمة كما ذكرنا هناك[1]، ومن ثم فإن أية احتمالات تخالف هذه القواعد تعد خطأً إملائيًّا في رسم المصحف الشريف، فيصبح من السهل اكتشاف مثل هذه الأخطاء.

ومن شأن تلك المعالجة لحرف الهمزة أن تمكّن من اكتشاف الأخطاء في رسم الهمزات في النص الشريف، ثم تدقيق النص آليًّا، والتأكيد والتصديق على صحة همزاته استهلاليةً أو متوسطةً أو متطرفة.

**ب ـ خصائص حركات الضبط بالشكل (حركات التشكيل):**

يمكن أن نعتمد في تدقيق حركات الضبط والشكل في القرآن الكريم على بعض القواعد اللغوية الخاصة بالضبط والتشكيل، وقد تحدثنا عن ذلك في مبحث تطوير آليات التعرف على الحروف العربية، إذ عرضنا لخصائص علامات الضبط العربية (الكسرة – الضمة – الفتحة – السكون – الشدة) وأشكالها ومواضعها من بعض الحروف[2]، ومن ثم يمكن للآلية المقترحة الاعتماد عليها في اكتشاف بعض الأخطاء في ضبط الشكل للنص الشريف، فأيّ مخالفة لأيٍّ من هذه القواعد تعني خطأً إملائيًّا في ضبط نص المصحف الشريف يمكن للآلية المقترحة –من خلال هذه القواعد– اكتشافه بسهولة.

وقد ذكرنا هناك أن لهذه الحركات خصائص إضافية – لاسيما من حيث الشكل – إذا هي دخلت على الحروف العربية، كل على حدة، وضربنا الهمزة مثالًا تطبيقيًّا على ذلك

---

١ – راجع: استخدامات الحروف العربية، سليمان فياض، ص ٢٤ .

٢ – راجع: نفسه، ص ١٦ .

من خلال ذكر احتمالات تشكيلها باختلاف موضعها من الكلمة على النحو الآتي:[1]

| الألف والهمزة | أول الكلمة | وسط الكلمة | آخر الكلمة | ملاحظات |
| --- | --- | --- | --- | --- |
| أ | تقبل من الحركات الفتحة والضمة | تقبل من الحركات الفتحة والضمة والسكون | تقبل من الحركات الفتحة والضمة والكسرة وصور التنوين كلها | |
| إ | الكسرة فقط | الكسرة فقط | الكسرة فقط | |
| ا | لا تضبط | لا تضبط | لا تضبط | |
| ء | تقبل من الحركات الفتحة | تقبل من الحركات الفتحة والضمة والتنوين بالفتحة مع إضافة الألف بعدها | تقبل من الحركات الفتحة والضمة والكسرة وتنوين هذه الحركات جميعا | معنى ذلك أن صور الهمزة مضبوطة هكذا: ءِ ، ءْ ، ءٌ لا تأتي وسط الكلمة. |
| ؤ | لا تأتي في بداية الكلمة | تقبل من الحركات الفتحة وا لضمة والسكون والتنوين بالفتحة مع إضافة الألف بعدها | تقبل من الحركات الفتحة و الضمة والكسرة وتنوين الضم والكسر | |
| ئ | لا تأتي في بداية الكلمة | تقبل من الحركات الفتحة وا لضمة والسكون والتنوين بالفتحة مع إضافة الألف بعدها | تقبل من الحركات الفتحة و الضمة والكسرة والسكون وتنوين الضم والكسر | |

١ – راجع: الإملاء والترقيم في الكتابة العربية، عبد العليم إبراهيم، ص ٤٥ وما بعدها.

| الألف والهمزة | أول الكلمة | وسط الكلمة | آخر الكلمة | ملاحظات |
|---|---|---|---|---|
| ى (ألف ممدودة) | لا تأتي في بداية الكلمة | تأتي في وسط الكلمة | لا تضبط | قبلها مفتوح دائما وصوره : (فتحة) َ+ ى – (تنوين بالفتحة) ً+ ى – (شدة وفتحة) َّ+ ى – (شدة وتنوين بالفتحة) ًّ+ ى |

جدول رقم (٢٦) لاحتمالات ضبط الهمزة والألف وشكلهما وفق مواضعهما المختلفة في الكلمة

فالجدول السابق يوضح احتمالات ضبط الهمزة والألف بالشكل وفق مواضعهما المختلفة، واختلاف شكل الهمزة نفسها باختلاف الحركات الداخلة عليها.

وهو ما فصلناه برسم العديد من الجداول التوضيحية في موضعه في مبحث تطوير التعرف على الحروف العربية[1].

١ – راجع: الإملاء والترقيم في الكتابة العربية، عبد العليم إبراهيم، ص ٥٤، ٥٥.

## ٣ - أدوات معالجة الرسم العثماني آليا:

إن من المهم بعد حوسبة القواعد العربية والإملائية وتغذية آلية التدقيق الإملائي الآلي بها؛ أن تتضمن هذه القواعد قواعد الرسم العثماني لضمان تغطية كافة الاستثناءات الإملائية والكتابية لكلمات القرآن الكريم، التي خالف فيها الرسم العثماني الكتابة العربية.

**قواعد الرسم العثماني:**

لرسم المصحف العثماني قواعد أحصاها علماء الرسم العثماني في مؤلفاتهم[١]، ويمكن إيجاز هذه القواعد في عدة أوجه على النحو الآتي:

**١ - الوجه الأول: الحذف**

- حذف الألف في قوله تعالى: ﴿ٱلْعَـٰلَمِينَ﴾ (الفاتحة:٢) حيث حُذفت الألف بعد العين، وقد كُتبت كذلك في جميع مواضعها في القرآن، والأصل في كتابتها حسب الرسم الإملائي (العالمين)[٢].




---

١ - من العلماء الذين أفردوا الرسم العثماني بالتأليف: الإمام أبو عمرو الداني، المتوفى سنة ٤٤٤هـ، في كتابه «الْمقنع في معرفة رسم مصاحف الأمصار، الإمام أبو محمد القاسم بن فِيرّه الشاطبي، صاحب حرز الأماني، المتوفى سنة ٥٩٠هـ، في كتابه «عقيلة أتراب القصائد في أسنى الْمقاصد، وهي نظم لكتاب الْمقنع الْمذكور، ولها شروح كثيرة. الإمام محمد بن إبراهيم الأموي الشهير بالخراز، المتوفى في أوائل القرن الثامن الهجري، في كتابه «مورد الظمآن»، وهو نظم بديع مشتمل على جل المسائل المذكورة في الكتب السابقة، وله شروحٌ، منها: دليل الحيران، للشيخ إبراهيم بن أحمد الْمارغني التونسي، المتوفى سنة ١٣٤٩ ه. الإمام أبو داود سليمان بن نجاح. المتوفى سنة ٤٩٦هـ، في كتابه «التنْزيل» .العلامة أبو عباس المراكشي،في كتابه «عنوان الدليل في رسوم خط التنزيل». العلامة محمد بن أحمد الشهير بالمتولي، في أرجوزته «اللؤلؤ المنظوم في ذكر جملة من المرسوم». العلامة محمد خلف الحسيني الذي شرح منظومته وذيل الشرح بكتاب أسماه: «مرشد الحيران إلى معرفة ما يجب اتباعه في رسم القرآن».

٢ - راجع: دليل الحيران شرح مورد الظمآن، إبراهيم بن أحمد المارغني التونسي، مكتبة الكليات الأزهرية - القاهرة - ١٩٨١م، ص ٤٦ وما بعدها.

- حذف الواو في قوله تعالى: ﴿وَٱلْغَاوُۥنَ﴾ (الشعراء:٩٤) وقد وردت في موضعين من القرآن، والأصل فيها ( الغاوون ) [١].

ٱلْغَاوُۥنَ
الرسم العثماني

الْغَاوونَ
الرسم الإملائي

- حذف الياء في قوله تعالى: ﴿ٱلنَّبِيِّـۧنَ﴾ (البقرة:٦١) وقد وردت كذلك في جميع مواضعها في القرآن، وعدد مواضعها ثلاثة عشر موضعاً، والأصل في كتابتها ( النبيين)[٢].

ٱلنَّبِيِّـۧنَ
الرسم العثماني

النَّبِيِّينَ
الرسم الإملائي

- حذف اللام في قوله تعالى: ﴿ٱلَّيْلِ﴾ (آل عمران:١٩٠) وقد كُتبت كذلك في جميع مواضعها، وعددها ثلاثة وسبعون موضعًا، والأصل فيها ( الليل ) [٣].

ٱلَّيْلِ
الرسم العثماني

اللَّيْل
الرسم الإملائي

- حذف النون في قوله تعالى: ﴿نُۨجِى﴾ (الأنبياء:٨٨) وهو الموضع الوحيد في القرآن، الذي حذفت فيه النون من ثلاثة مواضع وردت فيه الكلمة، والأصل في رسمها (ننجي ) [٤].

نُۨجِى
الرسم العثماني

نُنجِي
الرسم الإملائي

١ - راجع: نفسه، ص ٢٠٢-٢٠٣.

٢ - راجع: دليل الحيران شرح مورد الظمآن ، إبراهيم بن أحمد المارغني التونسي، ص ١٩٧-١٩٩.

٣ - راجع: نفسه، ص ٢٠٥-٢٠٧.

٤ - راجع: نفسه، ص ١٥٠-١٥١، وراجع: البرهان في علوم القرآن، محمد بن بهادر بن عبد الله الزركشي، تحقيق: محمد أبو الفضل إبراهيم، دار المعرفة - بيروت، د.ط ، ١٣٩١هـ، ١/ ٣٨٨-٤٠٨، والإتقان في علوم القرآن، عبد الرحمن بن أبي بكر، جلال الدين السيوطي، تحقيق: محمد أبو الفضل إبراهيم، دار التراث، د.ط، د.ت، ٤/ ١٤٧-١٥٠.

**٢- الوجه الثاني: الزيادة**

وتكون في الألف، والواو، والياء.

- الزيادة في الألف، قوله تعالى:﴿وَجِاْيٓءَ﴾ (الزمر:٦٩) وردت في موضعين، والأصل فيها ( وجيء )[١].

وَجِاْيٓءَ — الرسم العثماني

وَجِيءَ — الرسم الإملائي

- الزيادة في الواو، قوله تعالى: ﴿سَأُوْرِيكُمْ﴾ (الأعراف:١٤٥) وردت في موضعين، والأصل فيها (سأريكم)[٢].

سَأُوْرِيكُمْ — الرسم العثماني

سَأُرِيكمْ — الرسم الإملائي

- الزيادة في الياء، قوله تعالى: ﴿بِأَيْيْدٍ﴾ (الذاريات:٤٧) وهو الموضع الوحيد في القرآن، والأصل فيها (بأيد)[٣].

بِأَيْيْدٍ — الرسم العثماني

بِأَيْدٍ — الرسم الإملائي

**٣- الوجه الثالث: الهمز**

حيث وردت الهمزة في الرسم العثماني تارة برسم الألف، وتارة برسم الواو، وتارة برسم الياء.

---

١ - راجع: دليل الحيران شرح مورد الظمآن ، إبراهيم بن أحمد المارغني التونسي، ص ٢٤٢، ٢٤٥.

٢ - راجع: نفسه، ص ٢٥٩.

٣ - راجع: نفسه، ص ٢٥٦. وراجع: البرهان في علوم القرآن، محمد بن بهادر بن عبد الله الزركشي، (١/ ٣٨١-٣٨٨)، والإتقان في علوم القرآن، عبد الرحمن بن أبي بكر جلال الدين السيوطي، (٤/ ١٥١-١٥٢).

– ورودها ألفًا في قوله تعالى: ﴿لَتَنُوٓأُ﴾ (القصص:٧٦) وهو الموضع الوحيد، والأصل فيها ( لتنوء ) [١].

لَتَنُوٓأُ — الرسم العثماني

لَتَنُوء — الرسم الاملائي

– ورودها واوًا في قوله تعالى: ﴿يَبۡدَؤُاْ﴾ (يونس:٤) وهي كذلك في مواضعها الستة من القرآن، والأصل فيها ( يبدأ ) [٢].

يَبۡدَؤُاْ — الرسم العثماني

يَبْدَأ — الرسم الإملائي

– مجيئها ياءً في قوله تعالى: ﴿وَإِيتَآيِٕ﴾ (النحل:٩٠) وهو الموضع الوحيد من ثلاثة مواضع،والأصل فيها (وإيتاء)[٣]

وَإِيتَآيِٕ — الرسم العثماني

وَإِيتَاء — الرسم الإملائي

١ – راجع: دليل الحيران شرح مورد الظمآن، إبراهيم بن أحمد المارغني التونسي، ص ٢١٥.

٢ – راجع: نفسه، ص ٢٢٢–٢٢٣.

٣ – راجع: نفسه، ص ٢٢١، والإتقان في علوم القرآن عبد الرحمن بن أبي بكر، جلال الدين السيوطي، ٤/ ١٥٢–١٥٣

**٤- الوجه الرابع: البدل**

ويقع برسم الألف واوًا أو ياءً.

- مجيئها واوًا في قوله تعالى: ﴿ٱلصَّلَوٰةَ﴾ (البقرة:٣) وهي كذلك في جميع مواضعها الأربعة والستين، والأصل (الصلاة) ومثلها في ذلك ﴿ٱلزَّكَوٰةَ﴾ (البقرة:٤٣) والأصل (الزكاة) [١].

| الرسم العثماني | الرسم الإملائي |
| --- | --- |
| ٱلصَّلَوٰةَ<br>ٱلزَّكَوٰةَ | الصَّلاةَ<br>الزَّكاةَ |

- ومجيء رسمها ياءً في قوله تعالى: ﴿يَٰٓأَسَفَىٰ﴾ (يوسف:٨٤) والأصل فيها (يا أسفا) [٢].

| الرسم العثماني | الرسم الإملائي |
| --- | --- |
| يَٰٓأَسَفَىٰ | يَا أَسفَا |

ومن ذلك أيضاً، قوله تعالى: ﴿وَٱلضُّحَىٰ﴾ (الضحى:١) ولم ترد إلا في هذا الموضع، والأصل فيها (والضحا)[٣].

| الرسم العثماني | الرسم الإملائي |
| --- | --- |
| وَٱلضُّحَىٰ | وَالضُّحَا |

**٥- الوجه الخامس: الفصل والوصل**

فقد رُسمت بعض الكلمات في المصحف العثماني متصلة مع أنّ حقها الفصل،

١ - راجع: دليل الحيران شرح مورد الظمآن، إبراهيم بن أحمد المارغني التونسي، ص ٢٨٢-٢٨٣.

٢ - راجع: نفسه، ص ٢٦١.

٣ - راجع: نفسه، ص ٢٧٩-٢٨١.

ورُسمت كلمات أخرى منفصلة مع أنّ حقها الوصل .

- رسمتا متصلتين في مواضع، من ذلك قوله تعالى: ﴿بِئْسَمَا ٱشْتَرَوْا۟﴾ (البقرة: ٩٠) وهي كذلك في مواضعها الثلاثة.

| بِئْسَمَا | بِئْسَ مَا |
| --- | --- |
| الرسم العثماني | الرسم الإملائي |

والخلاصة أن حوسبة صور الكلمات القرآنية التي تختلف كتابتها في الرسم العثماني عنها في الرسم الكتابي إضافة إلى إحصاء عدد مرات تكرارها برسمها العثماني وبرسمها الكتابي في حالة وجوده في المصحف الشريف من خلال برامج الإحصاء والفهرسة اللغوية؛ تعد أحد أهم عناصر آلية التصحيح والتدقيق الآلي للنص الشريف، فمن الممكن بسهولة التأكد من دقة الرسم العثماني في المصحف الشريف من خلال مقارنة قواعد البيانات التي في الآلية والتي تخص النسخة الحاسوبية الأصلية للقرآن الكريم بمثيلاتها في النسخة المراد تصحيحها، ومن ثم اكتشاف الأخطاء في النسخ المراد تدقيقها وتصديقها.

## الخاتمة

عرضنا في هذا المبحث لمكونات آلية لغوية تمكننا من مراجعة النص القرآني كلمةً كلمةً، بل حرفًا حرفًا وحركةً حركةً، والتصديق على صحته، وتهدف إلى التدقيق الآلي؛ الإملائي واللغوي للنص القرآني في نسخه الرقمية برسمه العثماني. وأوضحنا كيف تعتمد الآلية في عملها على عدد من الأدوات الحاسوبية المقترحة، وهي:

الأدوات الإحصائية من خلال تطبيقات الإحصاء والفهرسة اللغوية.

الأدوات اللغوية من خلال عدد من الخصائص والقواعد اللغوية الحاسوبية.

أدوات معالجة الرسم العثماني آليًّا من خلال حوسبة قواعد الرسم العثماني وأوجه الاختلاف بينه وبين الرسم الكتابي للعربية

ويدعو المؤلف في كتابه إلى تبنّي الآلية وتوفير الدعم المؤسسي اللازم لها لأجل تنفيذها بوصفها خطوة أولى في طريق التدقيق الإملائي والنحوي للقرآن الكريم، كما يجدد الدعوة للباحثين لاستلهام النص الشريف بوصفه مادة خصبة للدراسات اللغوية الحاسوبية، التي يمكن من خلالها خدمة العلوم الإسلامية، ولاسيما القرآن الكريم والسنة النبوية المشرفة.



## المبحث الثاني

## تحليل النصوص باستخدام تقنيات الإحصاء اللغوية والاستفادة منها في صناعة مناهج تعليم العربية المعاصرة

### تمهيد:

يعمد هذا المبحث إلى تحليل مدونة لغوية للعربية باستخدام إحدى التقنيات اللغوية الحاسوبية؛ وهي تقنية الإحصاء والفهرسة اللغوية، بهدف المساعدة في تعليم العربية وتوفير مناهج عصرية لتعلم اللغة العربية المعاصرة، تكون مبنيّة على تحليل البيانات الضخمة لأجل التعليم والتعلم. فالمبحث يخرج بمجموعة من الإحصاءات الضرورية واللازمة في تعليم العربية المعاصرة للناطقين بها أو بغيرها، ومنها - على سبيل المثال - الخروج بإحصاء عن أكثر ١٠٠٠ كلمة تكرارًا في اللغة العربية المعاصرة، وكذلك أكثر ٥٠٠ متلازم وأكثر ١٠٠ مسكوك استخدامًا في العربية المعاصرة، وأخيرا أكثر ١٠٠ أداة نحوية استخدامًا في العربية؛ وهو أمر ذو بال في تعليم اللغة العربية للناطقين بها وبغيرها، وهذا سيساعد في تحديد الكلمات الأساسية لمتعلم العربية المعاصرة، ومن ثم يمكن للمعلمين وللمتعلمين التركيز على هذه الكلمات عند التعليم والتعلم، بل يجعل اللغويين ينطلقون من هذه الكلمات نحو معجم سريع وعصري للعربية المستخدمة أسوة بغيرها من اللغات التي سبقت في هذا المجال.

والمبحث الذي بين أيدينا يشرح الخطوات العملية للخروج بأمثال هذه النتائج السابقة وغيرها من خلال استخدام تقنيات الإحصاء والفهرسة اللغوية في معالجة

مدونة لغوية Corpus تضم مستويات لغوية مختلفة ومجالات دلالية مختلفة، حيث تؤدّي تقنيات الفهرسة اللغوية عملية التحليل الإحصائي لمفردات النصوص المعالجة، من إعداد فهارس لمفردات هذه النصوص مرتبة عدد مرات تكرارها، ومن ثم يمكن عرض هذه الفهارس في صورة قائمة عمودية مرتبة وفق الترتيب الهجائي التصاعدي أو التنازلي، أو وفق عدد تكرار الكلمات تصاعديًّا أو تنازليًّا، أو وفق طول الكلمة تصاعديًّا أو تنازليًّا. ومن أهم أمثلة هذه التطبيقات الإحصائية Concordance Version ٣,٢ [١]، و Textanz [٢].

١ - يمكن تحميله من خلال الموقع: http://www.concordancesoftware.co.uk/

٢ - يمكن تحميله من خلال الموقع: http://www.cro-code.com/textanz.jsp

## محتوى المبحث

مقدمة.
من الدراسات السابقة.
التعريف بالأدوات والتقنيات المستخدمة في الدراسة.
عرض تفصيلي لأهم تطبيقات الفهرسة الإحصائية
برنامج Concordance
برنامج Textanz (البرنامج المستخدم في المبحث).
مقدمة في التعريف بالمدونات اللغوية وأهدافها.
التعريف بالمدونة المستخدمة (مدونة لطيفة السليطي) في الدراسة:
خطوات معالجة المدونة اللغوية باستخدام تقنيات الفهرسة والإحصاء اللغوية.
البيانات التي يمكن استخراجها باستخدام تطبيقات الإحصاء الآلية (إجمالا):
إحصاء بأكثر ١٠٠٠ كلمة تكرارا في اللغة العربية المعاصرة.
إحصاء بأكثر ٥٠٠ متلازم استخداما في العربية في المعاصرة.
إحصاء بأكثر ١٠٠ مسكوك استخداما في العربية في المعاصرة.
إحصاء بأكثر ١٠٠ أداة نحوية استخداما في العربية.
- بيان كيفية الاستفادة من مخرجات تقنيات الإحصاء في مناهج تعليم العربية للناطقين بالعربية أو بغيرها.
- النتائج والتوصيات.

# مقدمة

لقد غدت الحاجة ملحّةً إلى أفكار جديدة في صناعة المناهج الدراسية؛ استنادا إلى الواقع المعيش والمواد المستخدمة، وذلك لتفعيل أثر هذه المناهج الدراسية في حياة كل المتعلمين والإفادة مما تعلموه في حياتهم، ومن ثم جاءت فكرة هذا المبحث بالبحث عن منهج معاصر للعربية يعتمد على المستخدَم من كلماتها وتراكيبها وقواعدها النحوية.

وليس أيسر من استخدام التقنيات الحديثة في اكتشاف هذه المناهج وصياغتها لتقديمها للدراسين، فيمكن من خلال استخدام تقنيات الإحصاء والفهرسة اللغوية الكشف عن أكثر الكلمات والمتلازمات والمسكوكات والأدوات النحوية استخدامًا في العربية المعاصرة من خلال الاعتماد على مدونة لغوية عربية معاصرة. وهو موضوع مطروح في عدد من اللغات الأجنبية ومنه على سبيل المثال – معجم كوبيلد الإنجليزي التعليمي – كما سيأتي في الدراسات السابقة.

## من الدراسات السابقة:

**مشروع معجم كوبيلد الإنجليزي التعليمي COBUILD English Dictionary:**[١]

مشـروع كبير، أنجزته مؤسسة كولينـز - كوبيلد Collins COBUILD الإنجليزية، المتخصصة في صناعة المعاجم اللغوية، ويهدف إلى وضع معجم تعليمي للغة الإنجليزية المعاصرة، يساعد متعلميها - من أهلها ومن غير أهلها - على فهم أعمق لتراكيبها واستعمالاتها اللغوية بعد حصر المفردات والتراكيب الشائعة، وبيان نسبة شيوع كل منها؛ كما اهتم صانعو المعجم بوضع أكبر عدد ممكن من الاستعمالات اللغوية لأقل عدد من المفردات، وهو ما يجعل تعلم اللغة والتمكن منها أمرًا ميسورًا. وقد نشـر معجم كوبيلد الإنجليزي التعليمي English Dictionary Helping Learners With Real English للمرة الأولى عام ١٩٨٧م، واعتمدت الطبعة الأولى على مدونة مكتوبة من ٢٠ مليون كلمة، جمعت من مصادر إنجليزية عديدة مما نشـر حتى الثمانينيات من القرن الماضي، وأعيد نشـر المعجم مرة أخرى عام ٢٠٠٠م، بإشراف المعجمي الإنجليزي - الإسكتلندي الأصل - جون سينكلير Jon Sinclair (١٩٣٣ - ٢٠٠٧م) - الذي كان أستاذًا للإنجليزية المعاصرة بجامعة برمنجهام حتى وافته المنية عام ٢٠٠٧م - بعد زيادة حجم مدونة المعجم إلى مائتي مليون (٢٠٠٠٠٠٠٠٠) كلمة[٢]. «ويعد معجم كولينز - كوبيلد واحدًا من أهم المعاجم الإنجليزية، يحظى بانتشار واسع في جميع أنحاء العالم؛ فلقد فطن صانعوه إلى المشكلات الأساسية في تعلم اللغة، وإلى أهمية الاستعمال اللغوي والوصف البراجماتي لمفردات اللغة، ووضعوا أسس وقوانين وإحصاءات وصلت بالمعجم إلى درجة عالية من الجودة والإتقان، فخرج كبير الفائدة، سهل الاستخدام، بكلمات شائعة، وتعبيرات واضحة، وقواعد نحوية ميسرة، واستعمالات لغوية منتقاة»[٣].

---

١ - راجع: مدونة معجم عربي معاصر، معالجة لغوية حاسوبية، المعتز بالله السعيد، ٢٠٠٧م. ص٦٩ وما بعدها.

2- Collins COBUILD York: English Dictionary Helping Learners With Real English. London. Introduction of Dictionary. Sincliar , J. and Others. (2000). P. viii.

٣ - راجع: مدونة معجم عربي معاصر، معالجة لغوية حاسوبية، المعتز بالله السعيد، ص٦٩ وما بعدها.

## التعريف بالأدوات والتقنيات المستخدمة في الدراسة:

### مقدمة في التعريف بتقنيات الإحصاء والفهرسة اللغوية[١]

تؤدّي تطبيقات الفهرسة الإحصائية عملية التحليل الإحصائي لمفردات النصوص المعالجة من خلال إعداد فهارس لمفردات هذه النصوص وتحديد عدد مرات تكرارها في هذه النصوص، وتعرض هذه الفهارس في صورة قائمة مرتبة وفق الترتيب الهجائي التصاعدي أو التنازلي، أو وفق عدد تكرار الكلمات تصاعديًّا أو تنازليًّا، أو وفقًا لطول الكلمة تصاعديًّا أو تنازليًّا[٢].

وتتيح هذه البرامج إمكانية حصر كلمات بعينها من خلال قائمةPick List ، أو استبعاد كلمات بعينها من خلال قائمة Stop List.

ومن أمثلة هذه التطبيقات Concordance Version ٢ , ٣[٣]، وTextanz[٤].

## عرض تفصيلي لأهم تطبيقات الفهرسة الإحصائية

### ١- برنامج Concordance[٥]

يدعم البرنامج اللغة العربية وحروفها، وتُدرج ملفات النصوص فيه بامتداد(.txt) فقط، حيث يقوم البرنامج بفهرسة الكلمات الموجودة بهذه النصوص، والكلمة عنده: مجموعة من الحروف المتصلِ بعضُها ببعض، الواقعةِ بين مسافتين قبلها وبعدها، ويعرض البرنامج النص في شكل قائمة واحدة من الكلمات (Headword) ويوضع أمام كل كلمة عدد تكرار هذه الكلمة في النص في قائمة (NO.)، ثم يعرض نسبة مئوية لهذا التكرار

---

١ – راجع: نفسه، ص٦٩ وما بعدها، وراجع أيضا: اللغة العربية المعاصرة في المجال السياسي، دراسة لغوية حاسوبية من واقع ذخيرة لغوية، هبة راشد، ص ٣٧، ٣٨ بتصرف.

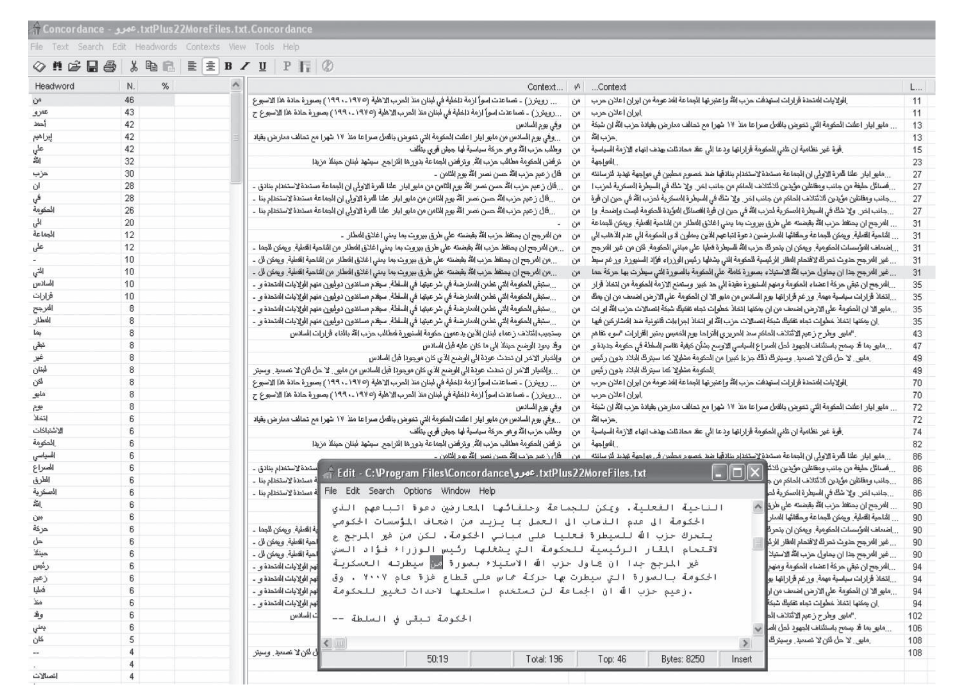
٢ – راجع: اللغة العربية المعاصرة في المجال السياسي، دراسة لغوية حاسوبية من واقع ذخيرة لغوية، هبة راشد، ص ٣٧، ٣٨ بتصرف. وراجع أيضا: مدونة معجم عربي معاصر، معالجة لغوية حاسوبية، المعتز بالله السعيد، ص٦٩ وما بعدها.

٣ – يمكن الحصول على البرنامج من خلال الرابط
http: //www.concordancesoftware.co.uk/ في: 2013-12-12.

٤ – يمكن الحصول على البرنامج من خلال الرابط
http: //www.cro-code.com /textanz.jsp في: 2013-12-12.

٥ – عرض الدكتور المعتز بالله السعيد لبرنامج Concordance في رسالته للماجستير، مدونة معجم عربي معاصر، معالجة لغوية حاسوبية، ص ٧٣، ٧٤.

بالمقارنة مع النص كله في قائمة (٪)، وذلك كما في الشكل:



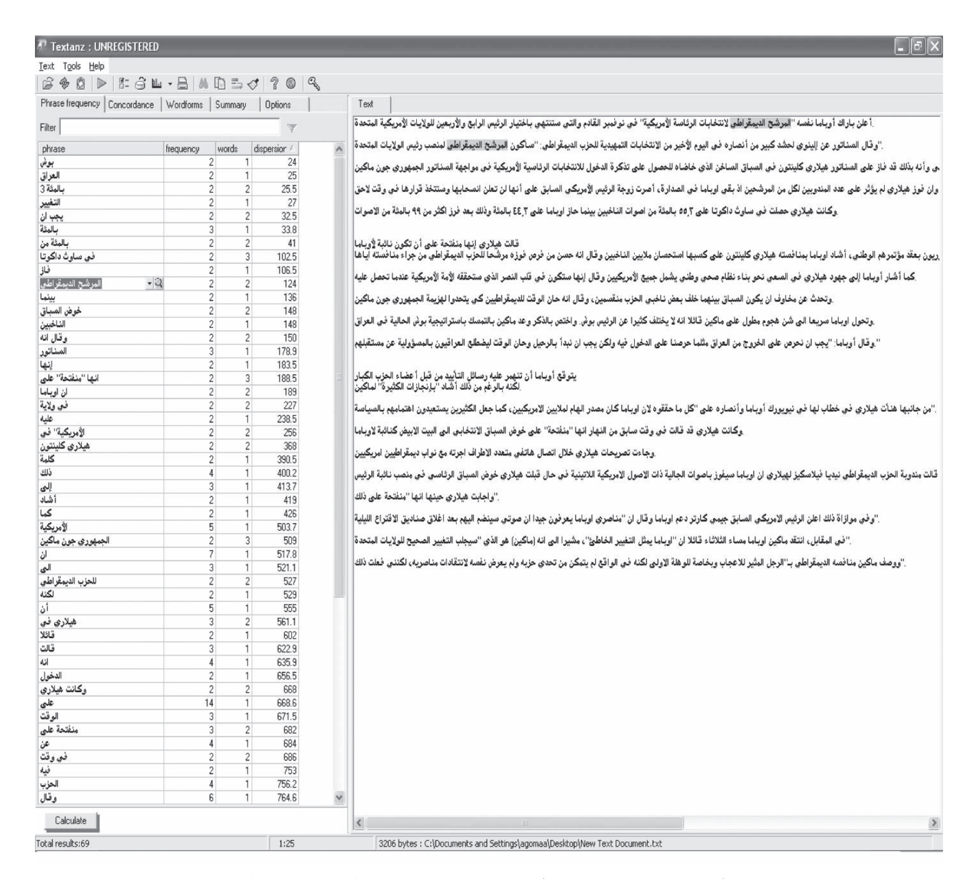

شكل رقم (٢١) لمعالجة برنامج Concordance للنصوص العربية

- وبالضغط على كل كلمة من كلمات القائمة على حدة، يظهر البرنامج النص الذي يحوي هذه الكلمة مع ١٠٠ كلمة حولها، ويمكن عمل قائمة (Stop list) لبعض الكلمات والأدوات المتكررة كحروف المعاني، ويمكن كذلك تخصيص بعض الكلمات فقط بالعد والفهرسة دون الأخرى من خلال إعداد قائمة (Pick List).

**٢- برنامج Textanz (البرنامج المستخدم في المبحث):**

يتميز برنامج Textanz عن برنامج Concordance بمعالجة التراكيب - إضافة إلى معالجة الكلمات - فباختيار وظيفة (Phrase)، يعرض البرنامج التراكيب

المتكررة في النص[1] مصحوبة بعدد مرات تكرارها، وعدد كلمات التركيب، فتركيب مثل: «المرشح الديمقراطي» عدد مرات وروده في النص المدرج (٢)، وعدد كلماته(٢).




شكل رقم (٢٢) لمعالجة برنامج Textanz للنصوص العربية

ويمكن للبرنامج تظليل الكلمات محل البحث في مواضعها في النص، ويتميز البرنامج بميزة أخرى وهي (Words forms) التي تختص بعرض الصور المختلفة للكلمة الواحدة، حيث يعرض للكلمات متشابهة الحروف في النص الواحد، فكلمات مثل: "الرئاسة"، "الرئاسية"، "الرئاسي" تأتي جميعًا تحت الوحدة "الرئاس"، ويمكن

١ – يمكن الإفادة من هذه الوظيفة في استخراج أكثر المتلازمات والتعابير المسكوكة شيوعًا في العربية المعاصرة، كما سيأتي.

التحكم في عدد حروف التشابه بتحديد أقصى عدد أو أقل عدد لحروف التشابه بين الكلمات.

ويمكن التحكم في نتائج البرنامج ومخرجاته عن طريق مجموعة من الخيارات؛ مثل عدد الكلمات في كل جملة، وعلامات تحديد بداية الجمل ونهايتها من خلال علامات الترقيم: ( ؟ )، ( ، )، ( ؛ )، ( ـ )، ( ! ). كما يظهر في الشكل الآتي:

شكل رقم (٢٣) لخيارات الفهرسة في برنامج Textanz

ويمكن للبرنامج تلخيص النص تلخيصًا يستوعب: عدد كلماته، ومتوسط كلمات الجملة الواحدة، ومتوسط كلمات الفقرة، ومتوسط حروف الكلمة الواحدة، وعدد الجمل، وعدد الفقرات، ومتوسط عدد الجمل للفقرة الواحدة، وعدد الأسطر، وعدد الحروف، وأطول كلمة، وأقصر كلمة، وأطول جملة وعدد حروفها، وأقصر جملة وعدد حروفها. ويمكن كذلك عمل قائمة بالكلمات ذات التكرار الكبير.

كلمات التوقف "Stop Words" : تضم كلمات مثل: «بين»، «عن»، «من»، «على»، «قبل»، «بعد»، «عبر»، «غير»، «مهما»، «ظل»، «هم»، «حول»، «دون»، «قد».

ومن الجدير بالذكر التأكيد على أن هذه التقنيات بحاجة إلى التطوير والمعالجة لضمان معالجة سليمة للعربية، فلاشك أن هناك مشكلات تخص العربية عند معالجتها بواسطة برامج الإحصاء اللغوي Concordance ، من أهمها[1]:

أ - مشكلة الصور المختلفة للدلالة الواحدة كما في الأعلام المترادفة - مشاكل الأخطاء الإملائية - السوابق واللواحق.

ب- مشكلة الصورة الواحدة للدلالات المختلفة (ظاهرة اللبس الحاسوبي الصرفي والدلالي)؛ اللبس الصرفي - اللبس الدلالي.

ج- مشكلة فصل التراكيب المتلازمة.

## مقدمة في التعريف بالمدونات اللغوية وأهدافها:

يعرّف د. المعتز بالله السعيد المدونة اللغوية بأنها كتلة غير منتظمة من النصوص المكتوبة أو المنطوقة التي تستخدم لدراسة جوانب اللغة، يمكن قراءتها والتعامل معها آليًّا بعد إدخالها على الحاسب الآلي، كما يمكن التحكم في بياناتها ومدخلاتها، بالإضافة أو الحذف أو التعديل من خلال قواعد بيانات (Databases) صُمِّمت للتعامل مع هذه النصوص. وتعدّ قاعدة البيانات الحاوية لنصوص المدونة اللغوية مخزنًا كبيرًا للغة، يرجع إليه وقت الحاجة، ويتحمل أي قدر من النصوص التي تضاف إلى المادة الأساسية مستقبلا[2].

---

١ - راجع للمزيد : من مشكلات التحليل النصي للمحتوى العربي على شبكة الإنترنت، د.سلوى السيد حمادة، - عمرو جمعة، المؤتمر الثامن للجمعية المصرية لهندسة اللغة (كلية الهندسة- جامعة عين شمس) - ٢٠٠٨.

٢ - مدونة معجم عربي معاصر، معالجة لغوية حاسوبية، المعتز بالله السعيد، ص ٤.

**التعريف بالمدونة المستخدمة (مدونة لطيفة السليطي) في الدراسة:**

عمد المؤلف إلى إحدى المدوّنات في العربية المعاصرة؛ «مدونة العربية المعاصرة» للباحثة لطيفة السليطي، التي يعود تاريخها إلى سنة ٢٠٠٤ م، ويبلغ عدد كلماتها نحو مليون كلمة، وتحديدًا «٨٤٢٦٨٤» كلمة، وعدد مقالاتها ٤١٥ مقالة، وأهم مجالاتها السياسة والاقتصاد والتعليم والرياضة، وهي مجمعة من الصحف والمجلات واللقاءات التليفزيونية والإذاعية وغيرها، ومن الممكن تحميل المدونة من الموقع الخاص بالباحثة لطيفة السليطي على موقع جامعة ليدز البريطانية[١].

| Name | Date modified | Type |
|---|---|---|
| Autobiography | 5/24/2014 3:26 PM | File folder |
| Children's Stories | 5/24/2014 3:26 PM | File folder |
| Economics | 5/24/2014 3:26 PM | File folder |
| Education | 5/24/2014 3:26 PM | File folder |
| Health and Medicine | 5/24/2014 3:26 PM | File folder |
| Interviews | 5/24/2014 3:26 PM | File folder |
| Politics | 5/24/2014 3:26 PM | File folder |
| Recipes | 5/24/2014 3:26 PM | File folder |
| Religion | 5/24/2014 3:26 PM | File folder |
| Science | 5/24/2014 3:26 PM | File folder |
| scienceb | 5/24/2014 3:26 PM | File folder |
| Short Stories | 5/24/2014 3:26 PM | File folder |
| Sociology | 5/24/2014 3:26 PM | File folder |
| Spoken | 5/24/2014 3:26 PM | File folder |
| Sports | 5/24/2014 3:26 PM | File folder |
| Tourist and Travel | 5/24/2014 3:26 PM | File folder |

شكل رقم (٢٤) لمجلدات مدونة لطيفة السليطي حسب مجالاتها

١ – راجع: مدونة العربية المعاصرة، لطيفة السليطي.
http: //www.comp.leeds.ac.uk /eric /latifa /research.htm في: ١٢-١٢-٢٠١٣.

## خطوات معالجة المدونة اللغوية باستخدام تقنيات الفهرسة والإحصاء اللغوية:

**١ - إعداد مادة المدونة للمعالجة الحاسوبية من خلال تحويل ملفاتها من صيغة (xml.) إلى صيغة (txt.) حتى يمكن لتقنيات الفهرسة الآلية التعرف عليها.**

| Name | Date modified | Type | Size |
|---|---|---|---|
| AUT01 | 8/17/2004 2:17 PM | XML Document | 39 KB |
| AUT02 | 8/17/2004 2:19 PM | XML Document | 16 KB |
| AUT03 | 8/17/2004 2:24 PM | XML Document | 9 KB |
| AUT04 | 8/17/2004 2:26 PM | XML Document | 14 KB |
| AUT05 | 6/27/2005 2:21 PM | XML Document | 17 KB |
| AUT06 | 8/17/2004 2:29 PM | XML Document | 15 KB |
| AUT07 | 6/27/2005 1:21 PM | XML Document | 31 KB |
| AUT08 | 8/17/2004 2:32 PM | XML Document | 17 KB |
| AUT09 | 8/17/2004 2:33 PM | XML Document | 12 KB |
| AUT10 | 6/27/2005 12:59 PM | XML Document | 22 KB |
| AUT11 | 8/17/2004 2:36 PM | XML Document | 14 KB |
| AUT12 | 8/17/2004 2:37 PM | XML Document | 9 KB |
| AUT13 | 8/17/2004 2:38 PM | XML Document | 14 KB |
| AUT14 | 8/17/2004 2:40 PM | XML Document | 17 KB |
| AUT15 | 8/17/2004 2:41 PM | XML Document | 33 KB |
| AUT16 | 8/17/2004 2:42 PM | XML Document | 31 KB |
| AUT17 | 8/17/2004 2:43 PM | XML Document | 18 KB |
| AUT18 | 8/17/2004 2:44 PM | XML Document | 47 KB |

شكل رقم (٢٥) لملفات أحد مجلدات مدونة لطيفة السليطي في صيغة XML

## ٢ – إدخال ملفات المدونة (في صيغة .txt) إلى المفهرس، وإعطاء الأمر ببدء عملية الفهرسة.

#conc - Notepad

File Edit Format View Help

Contemporary Arabic project, June, 2003   Whole speech of",2,43,1606.5
"المشكلات البيئية مثل التلوث بانواعه المختلفة، التصحر، استنزاف الموارد الطبيعية، واستنزاف الطاقة، الاسراف في استهلاك المياه والغذاء،
تزايد الحرارة، التوسع العمراني، قتل الغابات، الزيادة السكانية، الفقر، الامية، المخلفات بانواعها المختلفة، التغيرات المناخية، الظواهر الطبيعية
كالزلازل والبراكين والفيضانات والاعاصير، القاذورات، الحروب، الصيد الجائر، وغيرها",2,43,525674.5
"قول كبير باحثي قسم المخطوطات بسام بارود, وبين هذه المخطوطات قرابة خمسين مصحفا مخطوطا بخطوط متنوعة بين النسخ والمغربي
والفارسي, وأقدم مخطوطة في المجمع عنوانها (مجالس في التصوف) نسخت سنة 1277م وهي مجموع مقالات في الوعظ والإرشاد مجهولة
المؤلف",2,39,250150.5
"استقبلنا ببشاشة وجهه وبتواضعه الجم وحديثه الفياض باحترام الآخرين والعرفان بجميلهم, فقد حدثنا خلفان المهيري مدير الثقافة والفنون
مطولا عن كتاب (كنت شاهدا) الصادر حديثا عن المجمع لمؤلفه الذي كان محررا بمجلة العربي في بداية رحلتها, وفي",2,37,250150.5
"الدأب في البحث المستمر والدقيق وراء المعلومة ذات الأبعاد الاستراتيجية والتي تؤثر في الأوساط الثقافية والسعي الدائم خلف مجريات
الأحداث بالغة التأثير على العالم قاطبة وعلى البلاد العربية بشكل خاص وتحليلاتها لكي تتبلور نتائجها أمام أولي",2,36,250150.5
"مائتين وخمسين كتابا ودراسة في مختلف المجالات, ومائة وستين نشاطا علميا متنوعا ونشرات يومية ترصد وتحلل الأحداث المحلية والعربية
والعالمية توفر الرؤية الجلية لأصحاب القرار وتخدم السياسات الراهنة والمستقبلية للدولة استنادا إلى المعلومات الموثوقة والإحصاءات
الدقيقة",2,36,250150.5
"في واقع الحال لايمكن القول إن توزع المركز في الفلل السكنية عملي ولو كانت هذه الفلل متلاصقة أو متجاورة ومرتبطة تلفونيا وبشبكة
إنترنت, ولذلك عمل القائمون على المركز على تشييد بناء فخم يتناسب ومكانةالمركز",2,34,250150.5
"المجمع الثقافي أولى هذه المحطات وهو هيئة ثقافية عامة وقصر للحضارة أيضا, وعلى صعيد أساسي يؤدي دورا رائدا في نشر الثقافة وإثراء
الفكر وتشجيع الفنون الرفيعة والابتكارات العلمية وتأصيل الثقافات الوطنية والارتقاء بها",2,33,250150.5
"هذه العناوين أبحاث في مواضيع تاريخية وتراثية لها منزلتها عند الباحثين وفي تطور المجتمعات ولدى المهتمين ولذلك تتخاطفها المكتبات في

شكل رقم (٢٦) لأحد ملفات المدونة (في صيغة .txt)

## ٣ – الحصول على نسخة من مخرجات البرنامج في صيغة (.xlsx)

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Phrase | frequency | Words | Despersion |
| 2 | في | 22532 | 1 | 1336128.7 |
| 3 | على | 10428 | 1 | 1290747.3 |
| 4 | إلى | 7485 | 1 | 1375633.6 |
| 5 | عن | 4186 | 1 | 1318192.2 |
| 6 | ما | 3998 | 1 | 1252597.4 |
| 7 | هذه | 3426 | 1 | 1231669.9 |
| 8 | الذي | 3144 | 1 | 1370186.5 |
| 9 | التي | 2998 | 1 | 1246599.9 |
| 10 | كان | 2450 | 1 | 1404437.4 |
| 11 | لا | 2359 | 1 | 1204464.9 |
| 12 | ذلك | 2179 | 1 | 1267966.4 |
| 13 | و | 2165 | 1 | 950667.1 |
| 14 | بين | 2068 | 1 | 1267004.4 |
| 15 | كل | 1968 | 1 | 1315071.2 |
| 16 | هو | 1842 | 1 | 1227401.8 |
| 17 | لا | 1770 | 1 | 1182399.8 |
| 18 | بعد | 1758 | 1 | 1375670 |
| 19 | ان | 1713 | 1 | 999414.4 |

جدول رقم (٢٧) لمخرجات تطبيقات الإحصاء في صيغة (.xlsx)

**٤ - تحليل البيانات بعد فهرستها تحليلًا لغويًّا؛ وتصنيفها بغرض إعداد مناهج دراسية لتعليم العربية وتعلمها؛ لاستخراج البيانات الأساسية في عملية تعليم العربية المعاصرة، وستقتصر الدراسة على بعض النتائج على سبيل المثال، على النحو التالي:**

**البيانات التي يمكن استخراجها باستخدام تطبيقات الإحصاء الآليّة (إجمالًا):**

- إحصاء بأكثر ١٠٠٠ كلمة تكرارًا في اللغة العربية المعاصرة.
- إحصاء بأكثر ٥٠٠ متلازم استخدامًا في العربية في المعاصرة.
- إحصاء بأكثر ١٠٠ مسكوك استخدامًا في العربية في المعاصرة.
- إحصاء بأكثر ١٠٠ أداة نحوية استخدامًا في العربية.

وذلك على التفصيل الآتي:

**إحصاء بأكثر ١٠٠٠ كلمة تكرارًا في اللغة العربية المعاصرة:**

تبدأ مرحلة التحليل اللغوي بعد الحصول على مخرجات تطبيقات الفهرسة الإحصائية التي تتضمن عدد تكرار كل كلمة من كلمات المدونة بها مرتبة من الأعلى تكرارًا إلى الأقل تكرارًا؛ فإذا ما استثنينا الأدوات النحوية في العربية من الكلمات الأعلى تكرارًا وتردّدًا في المدونة اللغوية محل الدراسة، عن طريق استخدام قائمة (Stop List)؛ فيمكن حينئذ تحديد أكثر ١٠٠٠ كلمة تردّدًا وتكرارًا في العربية سوى هذه الأدوات النحوية، والتي يظهر بعضها من خلال الجدول الآتي:

| | A | B | C |
|---|---|---|---|
| 1 | Phrase | frequency | Words |
| 2 | الله | 1481 | 1 |
| 3 | العربية | 1345 | 1 |
| 4 | عام | 1227 | 1 |
| 5 | بعض | 1137 | 1 |
| 6 | العالم | 1133 | 1 |
| 7 | أكثر | 823 | 1 |
| 8 | السعودية | 816 | 1 |
| 9 | يمكن | 776 | 1 |
| 10 | يكون | 718 | 1 |
| 11 | العربي | 690 | 1 |
| 12 | الإسلام | 594 | 1 |
| 24 | المسلمين | 424 | 1 |
| 25 | سنة | 415 | 1 |
| 26 | العرب | 408 | 1 |
| 27 | الوقت | 408 | 1 |
| 28 | المرأة | 393 | 1 |
| 29 | خاصة | 388 | 1 |
| 30 | الإسلامي | 378 | 1 |
| 31 | بشكل | 375 | 1 |
| 32 | عبد | 375 | 1 |
| 33 | شيء | 372 | 1 |
| 34 | ذات | 363 | 1 |
| 35 | مجال | 353 | 1 |

| 13 | قال | 582 | 1 | 36 | يوم | 349 | 1 |
|---|---|---|---|---|---|---|---|
| 14 | الإنسان | 579 | 1 | 37 | جديدة | 347 | 1 |
| 15 | الناس | 549 | 1 | 38 | المملكة | 344 | 1 |
| 16 | آخر | 537 | 1 | 39 | العراق | 338 | 1 |
| 17 | العمل | 508 | 1 | 40 | يكن | 338 | 1 |
| 18 | عدد | 504 | 1 | 41 | وقال | 334 | 1 |
| 19 | الإسلامية | 474 | 1 | 42 | المجتمع | 331 | 1 |
| 20 | المتحدة | 461 | 1 | 43 | الأمريكية | 330 | 1 |
| 21 | تكون | 461 | 1 | 44 | القرن | 329 | 1 |
| 22 | الحياة | 450 | 1 | 45 | العالمية | 328 | 1 |
| 23 | نفسه | 425 | 1 | 46 | جميع | 324 | 1 |

جدول رقم (٢٨) لمخرجات تطبيقات الإحصاء في صيغة (.xlsx)

ويمكننا من خلال الجدول السابق التعرّف على بعض الكلمات الألف الأعلى تكرارًا وتردّدًا في المدونة محل الدراسة، حيث يحل لفظ الجلالة «الله» أولًا بتكرار قيمته ١٤٨١، وتحل كلمة «العربية» ثانية بتكرار قيمته ١٣٤٥ ، تليهما كلمة «عام» بتكرار قيمته ١٢٢٧ ... إلى آخرها.

**أما الكلمات الألف الأكثر تكرارًا في المدونة محل الدراسة فهي – على سبيل الحصر –:**

ابن – ارتفاع – ارد – استخدام – اسم – اسمه – الاتحاد – الاجتماعي – الاجتماعية – الاحتلال – الاختلاف – الاقتصاد – الاقتصادي – الاقتصادية – الاميركية – الاهتمام – الإبداع – الإسرائيلي – الإسلام – الإسلامي – الإسلامية – الإعلام – الإنتاج – الإنترنت – الإنجليزية – الإنسان – الإنسانية – الإيمان – الأبحاث – الأخير – الأخيرة – الأدب – الأدبية – الأساسية – الأسباب – الأسرة – الأسواق – الأسود – الأشجار – الأشياء – الأطباء – الأطفال – الأعمال – الأقليات – الأمة – الأمر – الأمر – الأمر – الأمراض – الأمريكي – الأمريكية – الأمور – الأمير – الأوربية – الأوسط – الأول – الأولى – الأيام – الآثار – الآخر – الآخرين – الآية – الباب – البحث – البحر – البحرية – البداية – البرامج – البرنامج – البشر – البشري – البشرية – البعض – البلاد – البنك – البيت – البيئة – البيئية – الت – التاريخ – التاريخية – التالي – التالية

– التجارة – التجارية – التجربة – التراث – التربية – الترجمة – التعامل – التعاون – التعبير – التعليم – التقدم – التقنية – التنمية – الثالث – الثاني – الثانية – الثقافة – الثقافي – الثقافية – الثورة – الجامعة – الجانب – الجديد – الجديدة – الجراحة – الجزء – الجسم – الجميع – الجهات – الجهاز – الحال – الحالات – الحالة – الحالي – الحجاب – الحداثة – الحديث – الحديثة – الحرارة – الحرب – الحركة – الحرية – الحصول – الحضارة – الحق – الحقيقة – الحكم – الحكومة – الحكومية – الحوار – الحي – الحياة – الحيوان – الحيوانات – الخارج – الخارجية – الخاص – الخاصة – الخدمات – الخلايا – الخليج – الخير – الداخل – الداخلية – الدراسات – الدراسة – الدعوة – الدفاع – الدكتور – الدم – الدنيا – الدواء – الدور – الدول – الدولة – الدولي – الدولية – الديني – الدينية – الذات – الرابع – الرأي – الرجل – الردة – الرسول – الرغم – الرواية – الرؤية – الرياض – الرئيس – الرئيسية – الزراعي – الزراعية – الزمان – الزمن – السابق – السابقة – السادات – الساعة – السبب – السعودي – السعودية – السفر – السلام – السلطة – السماء – السنة – السنوات – السوق – السؤال – السياحة – السياحي – السياحية – السيارات – السيارة – السياسة – السياسي – السياسية – السيد – الشاعر – الشباب – الشخصية – الشرعية – الشرق – الشرقية – الشركات – الشركة – الشريعة – الشعب – الشعبية – الشعر – الشعوب – الشمال – الشمس – الشهر – الشيخ – الشيعة – الصباح – الصحة – الصحراء – الصحية – الصراع – الصغير – الصغيرة – الصلاة – الصناعة – الصناعية – الصورة – الضوء – الطاقة – الطب – الطبية – الطبيعة – الطبيعي – الطبيعية – الطرق – الطريق – الطريقة – الطعام – الطفل – الطلاب – الطلب – الطويل – الطيران – الظاهرة – الظروف – العالم – العالمي – العالمية – العالي – العامة – العدد – العديد – العراق – العراقي – العراقية – العرب – العربي – العربية – العسل – العشرين – العصر – العقد – العقل – العلاج – العلاقات – العلاقة – العلم – العلماء – العلمي – العلمية – العلوم – العليا – العمر – العمل – العملية – العناصر – العين – الغرب – الغربي – الغربية – الغرفة – الفترة – الفرنسي – الفرنسية – الفضاء – الفقه – الفقهي – الفكر – الفلسطيني – الفلسطينية – الفن – الفنان – الفندق – الفنية – الفيلم – القاعدة – القانون – القاهرة – القدرة – القدس – القدم – القديم – القديمة – القراءة – القرار – القرآن – القرن – القسم

– القضية – القطاع – القلب – القول – القوى – القيام – القيم – الكاتب – الكائنات
– الكبرى – الكبير – الكبيرة – الكتاب – الكتابة – الكتب – الكثير – الكريم – الكلام
– الكويت – اللازمة – اللبنانية – اللجنة – اللغات – اللغة – اللقاء – الله – الليل –
المادة – الماضي – الماضية – المبيدات – المتحدة – المتخصصة – المثال – المجال – المجتمع
– المجتمعات – المجلس – المجمع – المحطة – المحكمة – المحلية – المختلفة – المدارس
– المدرسة – المدن – المدينة – المذاهب – المرأة – المرحلة – المركز – المريض – المزيد
– المسألة – المستقبل – المستوى – المسجد – المسلم – المسلمة – المسلمون – المسلمين
– المشاركة – المشروع – المشكلة – المصادر – المصري – المصرية – المعلومات – المغرب
– المكان – الملك – المملكة – المناسبة – المناطق – المنزل – المنطقة – المهمة – المواد
– الموارد – الموت – الموضوع – الموقف – المؤسسات – المياه – المئة – النار – الناس
– الناصر – النساء – النشاط – النشر – النص – النظام – النظر – النفس – النفسية
– النقد – النمو – النوع – النوم – الهدف – الهند – الهواء – الهيئة – الواحد – الواقع
– الوجه – الوجود – الوحدة – الوحيد – الوزراء – الوصول – الوضع – الوطن –
الوطني – الوطنية – الوقت – الولايات – اليهود – امرأة – امس – إبراهيم – إجراء –
إحدى – إدارة – إسرائيل – إسلامية – إضافة – إطار – إعادة – إقامة – إنتاج – إنسان
– إنشاء – أبناء – أبو – أجل – أحد – أحيانا – أخذ – أخرى – أخرى – أخرى – أريد
– أسباب – أشهر – أصبح – أصبحت – أصحاب – أعضاء – أعلى – أعمال – أفراد
– أفضل – أقل – أقول – أكبر – أكثر – ألف – ألف – أم – أمر – أمراض – أمريكا –
أمس – أمين – أنحاء – أنواع – أهل – أهمية – أوربا – أول – أيام – آثار – آخر – آل
– آلاف – باب – باسم – بالإضافة – بالرغم – بالفعل – بالنسبة – بداية – بدأ – بدأت
– برامج – برنامج – بسبب – بسرعة – بشكل – بصورة – بطريقة – بعدها – بعض –
بعضها – بعضهم – بعيد – بغداد – بقية – بلد – بلغ – بناء – بنسبة – بيت – بيروت
– تاريخ – تأثير – تبدأ – تبقى – تجد – تجربة – تحتاج – تحدث – تحديد – تحقق – تحقيق
– تحمل – تدخل – تر – ترى – تستطيع – تشكل – تصل – تطوير – تظهر – تعتبر
– تعتمد – تعد – تعرض – تعرف – تعمل – تعود – تعيش – تغيير – تقدم – تقديم –
تقع – تقول – تقوم – تكون – تل – تم – تماما – تماماً – تمثل – تنظيم – تنفيذ – توجد
– توفير – تؤدي – ثانية – ثلاث – ثلاثة – جاء – جاءت – جامعة – جان – جانب –

جائزة – جدا – جداً – جدة – جديد – جديدة – جزء – جعل – جميع – جميعا – جنوب – جهة – جهود – حاجة – حال – حال – حالات – حالة – حالياً – حجم – حدث – حدوث – حدود – حديث – حرارة – حرب – حركة – حرية – حسب – حسن – حسين – حقوق – حقيقة – حكم – حماية – حوالي – حياة – حياته – حياتي – حين – خارج – خاص – خاصة – خدمات – خدمة – خط – خطة – خلالها – خلف – خلق – خمس – خمسة – دُبَيْب – دار – دائما – دائماً – دخل – دراسة – درجة – دعم – دفع – دقيقة – دمشق – دور – دول – دولار – دولة – دين – ذات – ذاته – ذاتها – ذكر – ذل – رأس – رجل – رحلة – رد – رسالة – رسول – رضي – رمضان – روسيا – رؤية – رئيس – زمن – زوجها – زيادة – زيارة – سابك – ساعة – سبب – سبتمبر – سبق – سبيل – سطح – سعود – سعيد – سكان – سلامة – سن – سنة – سنوات – سوق – سوى – سؤال – سي – سيارة – سياسية – شاء – شاكر – شأن – شبكة – شخص – شخصية – شديد – شركات – شركة – شك – شكل – شمال – شهر – شيء – شيئا – شيئاً – صاحب – صحيح – صدام – صغيرة – صلى – صناعة – صوت – صورة – ضد – ضرورة – ضغط – ضمن – ضوء – طبيعة – طريق – طريقة – طلب – طوال – طويل – طويلة – ظاهرة – ظهر – ظهرت – ظهور – عادة – عالم – عالمية – عالية – عام – عاما – عاماً – عامة – عبارة – عبد – عبدالعزيز – عبدالله – عبر – عدة – عدد – عدم – عديدة – عربي – عربية – عرض – عشر – عشرة – عصر – عقد – علاج – علاقة – علم – علماء – علمية – عمر – عمره – عمل – عمله – عمليات – عملية – عهد – غرفة – غيره – غيرها – فتح – فترة – فرصة – فرنسا – فريق – فعل – فقال – فكرة – فلسطين – فلم – فما – فهم – قاعدة – قال – قالت – قام – قامت – قائلا – قائمة – قت – قدر – قدم – قرار – قسم – قصة – قضايا – قضية – قطاع – قطر – قطع – قطعة – قلت – قليل – قليلة – قوة – قوله – قيمة – كافة – كامل – كاملة – كبار – كبير – كبيرة – كتاب – كتابة – كتابه – كتب – كثير – كثيرا – كثيراً – كثيرة – كلمة – كله – كلها – كلية – كوبا – لبعض – لبنان – لجنة – لحظة – لغة – للعمل – لله – لمدة – مات – مادة – مائة – مباشرة – متر – متعددة – مثلا – مثلاً – مجال – مجالات – مجتمع – مجرد – مجلة – مجموعة – محاولة – محددة – محمود – مختلف – مختلفة – مدة – مدرسة – مدن – مدير – مدينة – مرات – مرة – مرحلة – مركز – مساحة – مسألة – مستوى

– مشروع – مشكلة – مصادر – مصر – مصر – معدل – معنى – معينة – مفهوم – مقابل – مقارنة – مكان – مكة – ملايين – ملعقة – مليار – مليون – مناطق – منتصف – منطقة – منع – مهما – مهمة – مواجهة – موسم – موضوع – مؤتمر – مؤسسة – مياه – ناحية – نتائج – نتيجة – نجد – نرى – نسبة – نشر – نص – نصف – نطاق – نظام – نظر – نفس – نفسه – نفسها – نفسي – نقص – نقطة – نقل – نهاية – نوع – هيكل – هيئة – واحد – واحدة – واسعة – واضحة – وأضاف – وبالتالي – وبعض – وجد – وجدت – وجه – وجود – وحدة – وحده – وحين – وخاصة – وراء – ورغم – وزارة – وزير – وسائل – وسط – وسلم – وصل – وضع – وعدم – وغير – وغيرها – وفق – وقال – وقت – وقع – وكل – ويقول – ويمكن – يأتي – يبدأ – يبدو – يبقى – يتعلق – يتم – يجب – يجب – يجد – يجعل – يجوز – يحتاج – يحدث – يحمل – يدخل – يريد – يزيد – يستطيع – يصبح – يعتبر – يعد – يعرف – يعمل – يعني – يعود – يعيش – يقدم – يقع – يقول – يقوم – يكن – يكون – يمثل – يمكن – ينبغي – يوجد – يوسف – يوم – يؤدي.

**إحصاء بأكثر ٥٠٠ متلازم استخدامًا في العربية في المعاصرة:**

**ينبغي بداية أن نعرّف بالمتلازمات اللفظية أو (التعابير السياقية)**

**المتلازم اللفظي Collocation:**

المتلازمات اللفظية: عبارة عن تلازم كلمتين أو أكثر بصورة شائعة في اللغة، تميل إلى ارتباط بعضها ببعض، لتماثل بين الملامح المعجمية لكل كلمة منها، ومثالها: «انتهك حُرْمة»، «خرق معاهدة»، «كلمة مأثورة»[١].

ويطلق عليها البعض أسماء أخرى مثل «المتصاحبات اللفظية» أو المتواردات أو المقترنات أو المترافقات اللفظية (collocating words) أو المصاحبات اللفظية أو التجمعات اللفظية[٢].

١ – التعابير الاصطلاحية والسِّياقية ومعجم عربي لها، د. علي القاسمي، بحث بمجلة اللسان العربي، مكتب تنسيق التعريب بالرباط، د.ت. ص٢٩.

٢ – نفسه، ص٢٩.

إذن هو تلازم لمفردتين أو أكثر في سياق مطرد، وليس في استعمال مسكوك، نحو: حرب ضروس، جريمة نكراء، خسارات جسيمة، أو هو مركب لفظي من كلمتين أو مجموعة من الكلمات التي تتوارد مع بعضها بشكل متلازم. فكلمة «هطلت» –على سبيل المثال– تقترن في أغلب الأحيان بكلمة «الأمطار» فنقول «هطلت الأمطار» ولا نقول «هطلت الرياح» مثلًا أو نقول «خفيف الظل» ولا نقول «سريع الظل» أو «هش الظل».

وقد تتكون هذه المتلازمات من: اسم وفعل مثل: «أدّى الزكاة» و»أقام الصلاة" و»عبر النهر» و»خاض المعركة». أو اسم واسم مثل: «حقن الدماء» و»أداء الدين» و»الوفاء بالعهد». أو صفة واسم مثل «عابس الوجه» و»ليّن العريكة» و»طويل الباع» أو اسم وحرف مثل «تخرج في جامعة ...».

**تحديد أكثر ٥٠٠ متلازم استخدامًا في العربية في المعاصرة:**

لتحديد المتلازمات ينبغي حصر النظر وتركيزه في مخرجات تطبيقات الإحصاء من التراكيب التي تتكون من كلمتين فأكثر من مخرجات برامج الفهرسة الإحصائية وتقسيمها حسب عدد كلمات كل تركيب، ابتداء بالتراكيب الصغرى ذوات الكلمتين ثم التراكيب ذوات الكلمات الثلاث وانتهاء بالتراكيب الكبرى ذوات الكلمات الأربع.

## فعلى مستوى التراكيب ذوات الكلمتين:

| | A | B | C |
|---|---|---|---|
| 1 | Phrase | frequency | Words |
| 2 | القرآن الكريم | 101 | 2 |
| 3 | الشرق الأوسط | 85 | 2 |
| 4 | حقوق الإنسان | 69 | 2 |
| 5 | اللغة العربية | 56 | 2 |
| 6 | العالم الإسلامي | 47 | 2 |
| 7 | دولار أمريكي | 44 | 2 |
| 8 | القرن التاسع | 41 | 2 |
| 9 | الخليج العربي | 40 | 2 |
| 10 | شهر رمضان | 38 | 2 |
| 11 | طه حسين | 38 | 2 |
| 12 | الاقتصاد السعودي | 32 | 2 |
| 13 | الصحة العالمية | 32 | 2 |
| 14 | الضفة الغربية | 32 | 2 |
| 15 | اللغة الإنجليزية | 31 | 2 |
| 16 | شبكة الإنترنت | 31 | 2 |
| 17 | حافظ سلامة | 30 | 2 |
| 18 | البث الفضائي | 26 | 2 |
| 19 | التراث العربي | 25 | 2 |
| 20 | الخطوط الجوية | 25 | 2 |
| 21 | حديثي الولادة | 25 | 2 |
| 22 | الحروب الصليبية | 23 | 2 |
| 23 | الحضارة الإسلامية | 23 | 2 |

جدول رقم (٢٩) لمخرجات تطبيقات الإحصاء من التراكيب ذوات الكلمتين

يُلحظ من الوهلة الأولى للنظر في مخرجات تطبيقات الفهرسة الإحصائية من التراكيب ثنائية الكلمات اشتمال هذه المخرجات على الأعلام والكيانات مثل: «القرآن الكريم» و»الشرق الأوسط»وهو أمر طبيعي؛ لتكرر مثل هذه الأعلام والكيانات في نصوص المدونة محل الدراسة، ومن الممكن الإفادة منه تعليميًّا.

أما التراكيب المتلازمة ثنائية الكلمات الأكثر تكرارًا مثل:»حديثي الولادة» و «تقنية

المعلومات» و «رجال الأعمال» فكلها تراكيب تتلازم في كلماتها عادة، ويمكن تقديمها لمتعلم العربية في صورها المتلازمة بوصفها أكثر المتلازمات تكرارًا وتردّدًا في العربية.

ومن ثم يمكن حصر أكثر المتلازمات الثنائية تكرارًا وتردّدًا في المدونة محل الدراسة فيما يأتي:

القرآن الكريم - الشرق الأوسط - حقوق الإنسان - اللغة العربية - العالم الإسلامي - دولار أمريكي - القرن التاسع - الخليج العربي - شهر رمضان - الاقتصاد السعودي - الصحة العالمية - الضفة الغربية - اللغة الإنجليزية - شبكة الإنترنت - البث الفضائي - التراث العربي - الخطوط الجوية - حديثي الولادة - الحروب الصليبية - الحضارة الإسلامية - الحضارة العربية - للأمم المتحدة - البحر الأحمر - القدس الشرقية - المدينة المنورة - مدينة القدس - والولايات المتحدة - الكائنات البحرية - الموارد البشرية - مليار ريال - الدين الإسلامي - الشرق والغرب - العرب والمسلمين - القرن الثامن - المياه الجوفية - أمير المؤمنين - تقنية المعلومات - الجزيرة العربية - مجلس الوزراء - الأمم المتحدة - البلاد العربية - التربية الفنية - الحياة الاجتماعية - الدول النامية - الشعب الفلسطيني - الاتحاد السوفييتي - الأدب الإسلامي - الأقلية المسلمة - الأمة الإسلامية - التجارة العالمية - التعليم الثانوي - الخدمات المصرفية - القراءة والكتابة - المحاصيل الزراعية - بالمنطقة الشرقية - بالولايات المتحدة - الاقتصاد العالمي - التنمية الاقتصادية - الدول الإسلامية - الرئيس المصري - السوق السوداء - السياحة الداخلية - الشرق الاوسط - حزب الله - رجال الأعمال - الدراسات العليا - الرئيس الفرنسي - القرن الخامس - حرب الخليج - دار الكتب - دول الخليج - البحر المتوسط - القرن السابع - المتحف العراقي - المجتمع المدني - دار العلوم - ريال سعودي - قطاع غزة - مجلس التعاون

| | A | B | C |
|---|---|---|---|
| 1 | Phrase | frequency | Words |
| 2 | المملكة العربية السعودية | 284 | 3 |
| 3 | الولايات المتحدة الأمريكية | 79 | 3 |
| 4 | رضي الله عنه | 61 | 3 |
| 5 | القرن التاسع عشر | 38 | 3 |
| 6 | الحرب العالمية الثانية | 28 | 3 |
| 7 | منطقة الشرق الأوسط | 25 | 3 |
| 8 | عمر بن الخطاب | 22 | 3 |
| 9 | الله- عز وجل | 18 | 3 |
| 10 | ثاني أكسيد الكربون | 18 | 3 |
| 11 | جميع أنحاء العالم | 17 | 3 |
| 12 | إن شاء الله | 16 | 3 |
| 13 | عليه الصلاة والسلام | 16 | 3 |
| 14 | القرن الثامن عشر | 14 | 3 |
| 15 | سوق حراء الدولي | 14 | 3 |
| 16 | شهر رمضان المبارك | 14 | 3 |
| 17 | مليار ريال سعودي | 14 | 3 |
| 18 | مليار دولار أمريكي | 13 | 3 |
| 19 | البنك السعودي الهولندي | 11 | 3 |
| 20 | دول مجلس التعاون | 11 | 3 |
| 21 | علي رضي الله | 11 | 3 |
| 22 | عمرو بن العاص | 11 | 3 |
| 23 | الامارات العربية المتحدة | 10 | 3 |

جدول رقم (٣٠) لمخرجات تطبيقات الإحصاء من التراكيب ذوات الكلمات الثلاث

حيث يمكن تحديد المتلازمات ثلاثية الكلمات الأكثر تكرارًا في المدونة المستخدمة على سبيل الحصر وفق ما يأتي:

المملكة العربية السعودية - الولايات المتحدة الأمريكية - رضي الله عنه - القرن التاسع عشر - الحرب العالمية الثانية - منطقة الشرق الأوسط - عمر بن الخطاب - الله- عز وجل - ثاني أكسيد الكربون - جميع أنحاء العالم - إن شاء الله - عليه الصلاة والسلام - القرن الثامن عشر - سوق حراء الدولي - شهر رمضان المبارك - مليار ريال سعودي - مليار دولار أمريكي - البنك السعودي الهولندي - دول مجلس التعاون - علي رضي الله - عمرو بن العاص - الإمارات العربية المتحدة - العالم العربي والإسلامي - الميزانية العامة الجديدة - دول مجلس التعاون - رضي الله عنهم - طلاب التعليم المدرسي - والولايات المتحدة الأمريكية - الرئيس جورج بوش - وزارة التربية والتعليم - أسلحة الدمار الشامل - الغرف التجارية الصناعية - أحداث ١١ سبتمبر - توت عنخ آمون - حسن كامل الصباح - منطقة الخليج العربي - القرن الخامس عشر - القرن السابع عشر - جائزة نوبل للسلام - القرآن الكريم والسنة - القرن الحادي والعشرين - القرن الواحد والعشرين - باللغتين العربية والإنجليزية - جامعة الدول العربية - جامعة الملك سعود - دول الخليج العربي - دول العالم الثالث - دول مجلس التعاون - عبدالله بن عبدالعزيز - الدول العربية والإسلامية.

وعلى صعيد التراكيب ذوات الكلمات الأربع:

| 1 | Phrase | frequency | Words |
|---|---|---|---|
| 2 | صلى الله عليه وسلم | 160 | 4 |
| 3 | التعليم المدرسي داخل المنزل | 39 | 4 |
| 4 | الخطوط الجوية العربية السعودية | 16 | 4 |
| 5 | مناطق المملكة العربية السعودية | 13 | 4 |
| 6 | مدن المملكة العربية السعودية | 12 | 4 |
| 7 | داخل المملكة العربية السعودية | 11 | 4 |
| 8 | علي رضي الله عنه | 10 | 4 |
| 9 | عمر بن عبد العزيز | 10 | 4 |
| 10 | خارج المملكة العربية السعودية | 9 | 4 |
| 11 | في كل أنحاء العالم | 9 | 4 |
| 12 | حقوق الإنسان في الإسلام | 8 | 4 |
| 13 | عمر - رضي الله عنه | 8 | 4 |
| 14 | غاز ثاني أكسيد الكربون | 8 | 4 |
| 15 | الاتحاد العربي في كوبا | 7 | 4 |
| 16 | البطالة في الاقتصاد السعودي | 7 | 4 |
| 17 | أنحاء المملكة العربية السعودية | 7 | 4 |
| 18 | بسم الله الرحمن الرحيم | 7 | 4 |
| 19 | دولة الإمارات العربية المتحدة | 7 | 4 |
| 20 | شركات اعادة التأمين العالمية | 7 | 4 |
| 21 | شركة فجر للاستثمار السياحي | 7 | 4 |
| 22 | شيخ الإسلام ابن تيمية | 7 | 4 |
| 23 | صندوق تنمية الموارد البشرية | 7 | 4 |

جدول رقم (٣١) لمخرجات تطبيقات الإحصاء من التراكيب ذوات الكلمات الأربع

يمكن تحديد المتلازمات ذوات الكلمات الأربع الأعلى تردّدًا وتكرارًا في المدونة على سبيل الحصر فيما يأتي:

صلى الله عليه وسلم – التعليم المدرسي داخل المنزل – الخطوط الجوية العربية السعودية – مناطق المملكة العربية السعودية – مدن المملكة العربية السعودية – داخل المملكة العربية السعودية – علي رضي الله عنه – عمر بن عبد العزيز – خارج المملكة العربية السعودية – في كل أنحاء العالم – حقوق الإنسان في الإسلام – عمر – رضي الله

عنه – غاز ثاني أكسيد الكربون – الاتحاد العربي في كوبا – البطالة في الاقتصاد السعودي – أنحاء المملكة العربية السعودية – بسم الله الرحمن الرحيم – دولة الإمارات العربية المتحدة – شركات إعادة التأمين العالمية – شركة فجر للاستثمار السياحي – شيخ الإسلام ابن تيمية – صندوق تنمية الموارد البشرية – الرسول عليه الصلاة والسلام – لا إله إلا الله – السعودية ودول الخليج العربي – القرن التاسع عشر الميلادي – القرن الخامس عشر الميلادي.

**إحصاء بأكثر ١٠٠ مسكوكٍ استخدامًا في العربية المعاصرة:**

قبل البدء في إحصاء أكثر المسكوكات استخدامًا في العربية المعاصرة سنعرّف بالمسكوكات أو التعابير الاصطلاحية.

المسكوك أو (التعبير الاصطلاحي) (Idiom): «وحدة لغوية (تضم أكثر من وحدة معجمية بسيطة)، يقع في الاستعمال اللغوي باطراد، وله دلالة ثابتة تختلف عن الدلالة المعجمية لمفرداته مفردة أو مركبة» [١]، فتكوين التعبير الاصطلاحي قائمٌ على «سلسلةٍ من الكلمات التي تُقيِّدها عوامل دلالية وتركيبية تجعل منها وحدة دلالية جديدة[٢]».

ولم يُولِ اللغويون القدماء اهتمامهم بهذا النوع من التعبيرات إلا فيما ندر، كما أن الباحثين المعاصرين لم يكونوا أكثر اهتمامًا من سابقيهم برصد هذه التعبيرات ودراستها؛ فلم تَحْظَ إلا باهتمام عدد قليل من الباحثين، انصرفت عنايتهم عادة إلى التعابير القديمة الموجودة في ثنايا كتب التراث"[٣]، ولم تكن هناك قواميس أحادية – أو حتى ثنائية – اللغة للمتلازمات في العربية على غرار قاموس Oxford Collocations في اللغة الإنجليزية مثلا، حتى أخرجت الدكتورة وفاء كامل فايد "معجم التعابير الاصطلاحية في العربية المعاصرة" عمدت فيه إلى حصر التعابير الاصطلاحية في العربية الحديثة والمعاصرة،

---

١ – راجع: علم الدلالة، د. أحمد مختار عمر، عالم الكتب، القاهرة، ط٥، ١٩٩٨م، ص٢٢–٢٣، وراجع: المعجم السياقي للتعبيرات الاصطلاحية، محمود إسماعيل صيني وآخرون، مكتبة لبنان ناشرون، بيروت، ط.١، ١٩٩٦م، ص ح (المقدمة).

٢ – الأسس النظرية لوضع معجم للمتلازمات اللفظية العربية، أسس المعجم النظري، د. محمد حلمي هليّل، مجلة المعجمية، تونس، ١٩٩٧م، عـ ١٢، ١٣، ص، ٢٢٦.

٣ – بعض صور التعبيرات الاصطلاحية في العربية المعاصرة، د. وفاء كامل، مجلة مجمع اللغة العربية بدمشق، المجلد ٧٨، الجزء٤، ص ٨٩٥–٨٩٦.

وهو معجم أحادي اللغة، مبني على مدوَّنة نصوص موسعة (Corpus based)، ويضم ما يقرب من ٤٠٠٠ تعبير اصطلاحي.

أما المدونة التي أمدّت المعجم بمادته فتضم نصوصًا لغوية تمثل اللغة العربية الحديثة والمعاصرة في مجالات مختلفة: [الأدب باختلاف أجناسه (من روايات وقصص ومسرحيات عربية ومترجمة للناشئة وللكبار) - الفنون - العلوم - الرياضة - الاجتماعيات - السياسة - الاقتصاد- الدوريات العلمية، والمجلات الأدبية والنقدية، والسياسية، والتربوية، والاجتماعية، والثقافية الفكرية، والنسائية، مما تصدر في كلٍّ من مصر والكويت والإمارات والسعودية وسوريا ولبنان ولندن... ] ويصل حجم المعجم إلى ٥٠٠ صفحة[(١)] .

وقسمت الدكتورة وفاء كامل التعابير الاصطلاحية المعاصرة تقسيمًا نحويًّا تركيبيًّا إلى[(٢)]:

المركب الإضافي، ومن أمثلته: آخِرُ العُنْقُودِ-نَسِيجُ المُجْتَمَع - أَبْنَاءُ الْبَطَّةِ السَّوْداء - ابْنُ الليْلِ- خَفِيفُ الظِّلِّ - تَحْتَ الشَّمْسِ- مَرْبِطُ الفَرَسِ - ابْنُ نُكْتَةٍ - ابْنُ حلالٍ - ابْنُ الحَلالِ - أَعْمَى الْقَلْبِ - نَظَافَةُ اليَد.

المركب الوصفي، ومن أمثلته: الطَّابُورُ الخَامِسُ - الْقَبْضَةُ الحَدِيدِيَّةُ - الْيَدُ الْعُلْيَا - الخُطُوطُ الْعَرِيضَةُ - السُّوقُ السَّوْدَاءُ - بابٌ خَلْفِيٌّ لِكَذا - بُرْجٌ عاجيّ - الْبَطَّةُ السَّوْداءُ - المَالُ الحَرَامُ - كَلامٌ فَارِغٌ - كَلَامٌ مُبَطَّنٌ - الكَلِمَةُ العُلْيَا- ابْتِسَامَةٌ صَفْرَاءُ.

المركب الفعلي، ومن أمثلته: أَثْلَجَ صَدْرَه - أَعْطَى الضَّوْءَ الأَخْضَرَ - يَرْفَعُ الرَّايَةَ الْبَيْضَاءَ - جَرَّ إِلَى الحَلْبَةِ - يَدْفِنُ الرُّؤُوسَ فِي الرِّمَالِ - ذَبَحَهُ بِسِكِّينٍ بَارِدٍ - يَصْطَادُ فِي المَاءِ الْعَكِرِ - أَرَاهُ النُّجُومَ ظُهْرًا - بَيَّضَ وَجْهَ فُلانٍ - أَخَذَ بِيَدِ فُلانٍ - أَخَذَ عَلَى خاطِرِهِ - أَدْخَلَ شَعْبَانَ فِي رَمَضَانَ - أَلْقَمَهُ حَجَرًا - نَسَفَ الجُسُورَ - نَشَرَ الغَسِيلَ القَذِرَ - نَفَخَ شِدْقَيْه - يَنْفُخُ فِي قِرْبَةٍ مَقْطُوعَةٍ - تَرَكَ الحَبْلَ عَلَى الْغَارِبِ.

المركب الاسمي، ومن أمثلته: اللَّعِبُ فِي الْوَقْتِ الضَّائِعِ - اللَّعِبُ عَلَى المَكْشُوفِ - اللَّعِبُ بِالنَّارِ - عَمَّال عَلَى بَطَّال - الخُرُوجُ مِنْ عُنُقِ الزُّجَاجَةِ - زَوْبَعَةٌ فِي فِنْجَان - رَبْط

١ - معجم التعابير الاصطلاحية في العربية المعاصرة، د. وفاء كامل، القاهرة، ط١، ٢٠٠٧، ص (ح).

٢ - بعض صور التعبيرات الاصطلاحية في العربية المعاصرة، د. وفاء كامل، مجلة مجمع اللغة العربية بدمشق، المجلد ٧٨، الجزء٤، ص ٨٩٨ وما بعدها.

الأَحْزِمَةِ عَلَى البُطُونِ – نَائِمٌ فِي الْعَسَلِ – فلانٌ نَفَسُهُ قَصِيرٌ – بصلته محروقة.

المركب العباري، ويتكون من شبه جملة، ومن أمثلته: بالفم المليان– عَلَى كَفِّ عفْرِيت – بَعْدَ خَرَابِ مَالطَة – فَوْقَ صَفِيحٍ سَاخِنٍ – بِالْبَرَكَة – في ذيل القائمة – (أَصْبَحَ) فِي خَبَرِ كَانَ – بأُمِّ أُذُنِي – بِالبَاعِ والذِّرَاعِ – بِدَمٍ بَارِدٍ – بربطة المعلم– بدون لفّ أو دوران.

ومن المهم إيجاز الفروق بين التعابير الاصطلاحية والتعابير السياقية كما نص عليها د. علي القاسمي فيما يلي[1]:

إمكان استشفاف معنى التعبير السياقي من فهم دلالة عناصره، أو الكلمات المكوّنة له على حدة، فقولنا: «خرق المعاهدة» يعني «انتهك الاتفاقية»؛ لأنّ «خرق» تعني «انتهك»، و»المعاهدة» تعني «الاتفاقية».

يمكن ذكر أحد أجزاء التعبير السياقي دون ذكر الجزء الثاني المكوّن له، فنستطيع أن نقول: «القدس» دون أن نردفها بكلمة «الشريف».

تمتاز التعابير السياقية بقدرتها على التنوع أي إنّه يمكن تبديل الكلمات المكوّنة للتعبير السياقي بكلماتٍ مماثلةٍ لها دلالتها دون الإخلال بمعنى التعبير الكلي، ففي التعبير السياقي «ثلةٌ من الجيش» يمكن أن نقول: «جماعةٌ من الجنود»، و»مجموعةٌ من العسكر».

١ – راجع: التعابير الاصطلاحية والسِّياقية ومعجم عربي لها، د. علي القاسمي، بحث بمجلة اللسان العربي، مكتب تنسيق التعريب بالرباط، د.ت. ص٢٩–٣٠.

### تحديد المسكوكات في المدونة محل الدراسة بواسطة تطبيقات الفهرسة الآلية:

| | A | B | C |
|---|---|---|---|
| 1 | Phrase | frequency | Idiom Words |
| 2 | سبيل الله | 18 | 2 |
| 3 | في سبيل الله | 17 | 3 |
| 4 | بغض النظر عن | 14 | 3 |
| 5 | وعلى سبيل المثال | 13 | 3 |
| 6 | على أيدي | 13 | 2 |
| 7 | على يديه | 10 | 2 |
| 8 | العلاقات العامة | 6 | 2 |
| 9 | وصلة الأرحام | 6 | 2 |
| 10 | على مدار الساعة | 5 | 3 |
| 11 | على مدار العام | 5 | 3 |
| 12 | ابن آدم | 4 | 2 |
| 13 | بعد العمل الجراحي | 3 | 3 |
| 14 | على قلب رجل واحد | 2 | 4 |
| 15 | النسيج الاجتماعي | 2 | 2 |

جدول رقم (٣٢) لمخرجات تطبيقات الإحصاء من بعض المسكوكات

تتعدد المسكوكات في العربية بتعدد عدد كلماتها، فهناك المسكوكات ثنائية الكلمات وهناك المسكوكات ثلاثية الكلمات، إلى آخر أنواعها المتعددة بتعدد كلماتها، وتتفاوت أعدادها وتردداتها وتكراراتها في العربية المعاصرة حسب أعداد كلماتها؛ فالدراسة الإحصائية تنبئ عن أن أعداد المسكوكات الثنائية وتكرارها وترددها في العربية أكثر من أعداد المسكوكات الثلاثية، وأعداد المسكوكات الثلاثية أكثر من أعداد الرباعية بطبيعة الحال.

## إحصاء بأكثر ١٠٠ أداة نحوية استخدامًا في العربية:

**إن الملحوظة الأولى في عملية التحليل اللغوية للمدونة محل الدراسة تكمن في أن الكلمات الأعلى تكرارًا وتردّدًا في المدونة من نصيب الأدوات النحوية في العربية:**

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Phrase | frequency | Words | Despersion |
| 2 | في | 22532 | 1 | 1336128.7 |
| 3 | على | 10428 | 1 | 1290747.3 |
| 4 | إلى | 7485 | 1 | 1375633.6 |
| 5 | عن | 4186 | 1 | 1318192.2 |
| 6 | ما | 3998 | 1 | 1252597.4 |
| 7 | هذه | 3426 | 1 | 1231669.9 |
| 8 | الذي | 3144 | 1 | 1370186.5 |
| 9 | التي | 2998 | 1 | 1246599.9 |
| 10 | كان | 2450 | 1 | 1404437.4 |
| 11 | لا | 2359 | 1 | 1204464.9 |
| 12 | ذلك | 2179 | 1 | 1267966.4 |
| 13 | و | 2165 | 1 | 950667.1 |
| 14 | بين | 2068 | 1 | 1267004.4 |
| 15 | كل | 1968 | 1 | 1315071.2 |
| 16 | هو | 1842 | 1 | 1227401.8 |
| 17 | لا | 1770 | 1 | 1182399.8 |
| 18 | بعد | 1758 | 1 | 1375670 |
| 19 | ان | 1713 | 1 | 999414.4 |

جدول رقم (٣٣) لمخرجات تطبيقات الإحصاء من بعض الأدوات النحوية الأكثر تكرارًا في العربية

فحرف الجر «في» هو الأداة الأعلى تكرارًا بين أدوات النحو العربية، بتكرار قيمته ٢٢٥٣٢ ، يليه حرف الجر «على» بتكرار قيمته ١٠٤٢٨ ، يليه حرف الجر «إلى» بتكرار قيمته ٧٤٨٥ ، وهكذا إلى آخر هذه الأدوات. وهو تحليل يشي بأهمية حروف الجر ودورها في ترابط الجملة العربية ومن ثم اعتبارها أَوْلى الدروس العربية بالتعلم لدارس العربية. وفي هذا إشارة إلى استخدام الفعل في العربية بشكل كبير، فحروف الجر في العربية تتعلق بالأفعال، ومن ثم يمكن للمعلم والمتعلم أن يستنتج أهمية الفعل في العربية وشيوع الجملة الفعلية بشكل أكبر من الجملة الاسمية.

ويلي حروفَ الجر من الأدوات أسماءُ الإشارة؛ فاسم الإشارة (هذه) يتكرر في المدونة ٣٤٢٦ مرة ، فيما يتكرر اسم الإشارة (ذلك) ٢١٧٩، وهو مؤشر على المنزلة الثانية للدروس النحوية عامة أو للأدوات النحوية خاصة بعد حروف الجر.

ثم يأتي دور الأسماء الموصولة بوصفها أكثر الكلمات تكرارًا في العربية؛ فيتكرر الاسم الموصول «الذي» ٣١٤٤ مرة ، يليه الاسم الموصول «التي» بتكرار قيمته ٢٩٩٨.

يلي ذلك الفعل الناسخ «كان» بتكرار قيمته ٢٤٥٠ في المدونة محل الدراسة، ثم حرف النفي «لا» بتكرار قيمته ٢٣٥٩.

### ويمكن حصر هذه الأدوات المائة فيما يأتي:

في – على – إلى – عن – ما – هذه – الذي – التي – كان – لا – ذلك – و – بين – كل – هو – لا – بعد – ان – كما – حتى – قد – وقد – كانت – التي – وفي – ولا – غير – وهو – ثم – أو – عليه – فيها – قبل – مع – بها – فيه – تلك – به – الذين – فى – لها – بل – مثل – هناك – ولكن – وهي – عند – أو – وكان – خلال – منذ – عليها – و – فقد – الآن – دون – فإن – مع – إن – علي – أو – هل – حيث – عندما – وما – أو – ولم – إن – أو – بن – لقد – لكن – حول – هي – ليس – التي – فقط – فيما – إليه – بما – أيضا – داخل – خلال – وذلك – مع – عنه – والتي – لم – تحت – له – وعلى – إنه – أمام – لم – حيث – إلا – إليها – إلا.

ولأن من سمات العربية ارتباط كلماتها بالسوابق واللواحق؛ فقد أتت هذه الكلمات مكررة ضمن المائة الأولى من الأدوات النحوية الأكثر تكرارًا في العربية لارتباطها بسوابق ولواحق مختلفة ولعدم اعتماد تطبيقات الفهرسة الإحصائية على التحليل الصرفي عند الإحصاء، وهي من مشكلات التطبيقات اللغوية العربية.[1]

– بيان كيفية الإفادة من مخرجات تقنيات الإحصاء في مناهج تعليم العربية للناطقين بالعربية أو لغير الناطقين بها:

ينبغي التأكيد على أنه ليس الهدف من المبحث اكتشاف أكثر الكلمات أو المتلازمات أو المسكوكات أو الأدوات النحوية دورانًا في العربية المعاصرة، بل الهدف هو رسم منهج

١ – راجع: بحث «من مشكلات التحليل النصي للمحتوى العربي على شبكة الإنترنت» د.سلوى السيد حمادة، عمرو جمعة، مؤتمر هندسة اللغة ٢٠٠٨.

محدد وطريقة علمية صحيحة لاكتشاف هذه الكلمات أو المتلازمات أو المسكوكات أو الأدوات النحوية الأكثر دورانًا.

فالمبحث إنما يرسم الطريق ويضع المنهج، ومن الممكن لباحث آخر أن يعمد إلى مدونة أكبر من مدونة المبحث تتكون من ملايين الكلمات، فتكون النتائج أدق وأكثر شمولًا للعربية المعاصرة كلها، وأشمل فائدة لمتعلم العربية ومعلمها. ومن الممكن كذلك استخدام المنهج ذاته لمعرفة أكثر الكلمات دورانًا في نصوص تخصص ما، كأن يعمد الباحث إلى مدونة متخصصة ويحللها بواسطة تطبيقات الفهرسة الآلية لمعرفة أكثر الكلمات دورانًا في نصوص هذا التخصص، فيكون التركيز عليها عند تعلم كلمات هذا الحقل الدلالي المتخصص.

ومن المهم التنبيه إلى أن ذلك لا يقلل من قيمة المبحث ونتائجه، وهي نتائج محددة وفقًا للمدونة مادة الدراسة، ولكن الدراسة تقدم تصوّرًا مبدئيًّا لما يمكن أن تكون عليه النتائج في حال معالجة مدونة لغوية أكبر حجمًا من المعالجة في هذه الدراسة.

ويمكن إيجاز كيفية الإفادة من مخرجات تقنيات الإحصاء في مناهج تعليم العربية للناطقين بالعربية أو بغيرها، كل على حدة.

ففيما يخص إحصاء أكثر ١٠٠٠ كلمة تكرارًا في اللغة العربية المعاصرة يمكن الإفادة من هذه النتائج في تحديد أكثر الكلمات دورانًا وتكرارًا في العربية المعاصرة؛ كما يمكن:

ربط هذه الكلمات حسب حقولها الدلالية وتقديمها لدارس العربية ومتعلمها من الناطقين بها أو بغيرها ضمن الحقل اللغوي الخاص بها، فالكلمات: ألف – عشر – مائة – ثلاث – ... كلمات تدخل تحت الحقل الدلالي «الأرقام والأعداد» ، والكلمات: يوم – سنة – عام – شهر – أسبوع – ... كلمات تحت الحقل الدلالي «التاريخ والوقت».

ربط هذه الكلمات حسب جذورها الثلاثية والرباعية في العربية لتسهيل ربط الكلمات ذوات الجذر الواحد، فالكلمات: "عمل" – "عملُه" – "عمليات" – "عملية" ذات جذر واحد؛ هو "عمل"، وكذلك الكلمات: "الدول" – "الدولة" – "الدولي" – "الدولية" ذات جذر واحد؛ جذرها "دول". وهي خطوة نحو معجم لغوي لأكثر الكلمات تكرارًا في العربية المعاصرة أسوة بهذا النوع من المعاجم في بعض اللغات.

تصنيف هذه الكلمات حسب أنواع الكلم في العربية إلى "الاسم" و"الفعل" مع تجاهل تصنيف "الحرف" لوجوده ضمن قائمة الأدوات النحوية الأكثر استعمالًا في العربية المعاصرة.

أما ما يتعلق بإحصاء أكثر ٥٠٠ متلازم استخدامًا في العربية في المعاصرة، وإحصاء أكثر ١٠٠ مسكوك استخدامًا في العربية في المعاصرة فيمكن الإفادة منهما في الخروج عن النمط المألوف في المعاجم العربية إلى معاجم للمتلازمات والمسكوكات، وهو فرع تحتاج المكتبة العربية فيه إلى العديد من المعاجم المتخصصة، مع إمكان تقديم ترجمة لهذه المتلازمات والمسكوكات إلى بعض اللغات، وهذا سيساعد الكثير من المترجمين -إضافة لمتعلمي العربية من الناطقين بغيرها - في التعرف على معاني هذه المتلازمات والمسكوكات في العربية بلغاتهم الأم. فترجمة المسكوك تختلف بشكل جذري عن ترجمة ما سواه في أي لغة.

أما إحصاء أكثر ١٠٠ أداة نحوية استخدامًا في العربية، فيُرجى منها إحصاء أكثر مائة أداة نحوية استخدامًا في العربية المعاصرة، وتحديدها وربطها ببعضها وتحديد دروسها النحوية الكلية، لتقديمها لمتعلمي العربية المعاصرة من الناطقين بها أو بغيرها في الإطار النحوي الكلي بشكل عام. فالأدوات المائة الأكثر استخدامًا في العربية تجمعها - من بين ما تجمعها - حروف الجر وأسماء الإشارة والأسماء الموصولة والضمائر والظروف وأدوات النفي والاستثناء، وهي دروس نحوية من المهم لدارس العربية تعلمها كأساس لتعلم نحوِ العربية المعاصرة؛ أعني نحوها المعتمد على الواقع المستخدَم وليس على التنظير البعيد عن الواقع أو ما يمكن أن نسميه بالمهمل من اللغة وقواعدها.

## النتائج والتوصيات:

الخروج بإحصاءات عن أكثر ١٠٠٠ كلمة تكرارًا في اللغة العربية المعاصرة، وعن أكثر ٥٠٠ متلازمٍ وأكثر ١٠٠ مسكوكٍ استخدامًا في العربية المعاصرة، وأخيرًا أكثر ١٠٠ أداة نحوية استخدامًا في العربية.

رسم هذا المبحث منهجًا محدّدًا وطريقة علمية لاكتشاف الكلمات أو المتلازمات أو المسكوكات أو الأدوات النحوية الأكثر دورانًا في العربية المعاصرة.

يوصي المؤلف بمعالجة مدونة لغوية أكبر حجمًا، تتكون من ملايين الكلمات تكون شاملة لمستويات العربية المختلفة لضمان نتائج دقيقة وتفصيلية.

يوصي المؤلف باتخاذ نتائجه من الكلمات والمسكوكات والمتلازمات والأدوات النحوية نواة لمعجم عربي معاصر وقاموس متعدد اللغات من العربية إلى غيرها من اللغات.

# الخاتمة

حاول الكتاب فيما مضى من صفحاته أن يجعل من التقنيات اللغوية الحاسوبية محورًا للدراسة، فعرض للمعايير اللغوية لتقييم بعض التقنيات الحاسوبية التي تعالج اللغة العربية؛ من محركات البحث، إلى برامج التدقيق الإملائي والنحوي، مرورًا ببرامج التعرف الضوئي على الحروف العربية.

وتوصل الكتاب إلى تحديد خمسة معايير لغوية لتقييم تعامل محرك البحث مع كلمات البحث على المستوى الصرفي، وإلى تحديد ثلاثة معايير لغوية لتقييم تعامل محرك البحث مع كلمات البحث على المستوى الدلالي. كما توصل إلى تسعة معايير لتقييم برامج التعرف الضوئي على الحروف العربية وعلى حركات الضبط بالشكل. وتوصل الكتاب كذلك إلى سبعة معايير لغوية لتقييم برامج التدقيق الإملائي والنحوي.

ثم انطلق ليقدم رؤى تطويرية لتقنيتين من التقنيات اللغوية؛ أولاهما برامج التعرف الآلي على الحروف العربية؛ إذ اقترح المؤلف نموذجًا لغويًّا مدمجًا لحل مشكلات التعرف على الحروف العربية على مستويي النصوص المعالَجة؛ مستوى النصوص غير المضبوطة بالشكل، ويقوم على خمس خطوات من خطوات المعالجة المرتبة التي تقود فيه كل خطوة إلى أختها. ومستوى النصوص المضبوطة بالشكل اعتمادًا على القواعد اللغوية المحوسبة، ضاربًا المثال بحرف الهمزة.

وثانية الرؤى التطويرية تخص برامج التدقيق اللغوي؛ إذ اقترح الكتاب فكرة آلية لغوية هدفها تقريب العربية من أهلها، باستخدام أداة لغوية مدمجة (Plugin) تعتمد على مخرجات برامج التدقيق اللغوي، وعلى قاعدة بيانات تحليلية تمكن الآلية من تقديم التأصيل العلمي للقضية محل الإشكال إملائيًّا أو نحويًّا أو تركيبيًّا بغرض تعليمي.

ثم أراد المؤلف أن يسلك بكتابه آفاقًا جديدة لاستخدام بعض التقنيات اللغوية فاقترح مراجعة النص القرآني وتدقيقه آليًّا من خلال آلية لغوية هدفها التدقيق الإملائي واللغوي للنص القرآني في نسخه الرقمية برسمه العثماني، معتمدة في عملها على ثلاث من الأدوات الحاسوبية؛ الإحصائية (على مستوى الحرف والكلمة ونهايات الآي وصفحات المصحف) واللغوية (من خلال خصائص الحروف العربية وخصائص حركات الضبط بالشكل وخصائص الحركات العربية مع بعض الحروف (تطبيقا على الهمزة)، وأخيرًا أدوات معالجة الرسم العثماني آليًّا من خلال حوسبة قواعد الرسم العثماني (حذف بعض الحروف وزيادة بعض الحروف وكتابة الهمزة على غير القياس واستبدال حرف بحرف والفصل والوصل بين أجزاء كلمات القرآن).

وأخيرًا عرض المؤلف لاستخدام تقنيات الإحصاء اللغوية في تحليل النصوص اللغوية بهدف الخروج بإحصاءات دقيقة من شأنها إفادة المستخدم العربي والتأثير في واقعه، وقد خرج الكتاب بعد معالجته لمدونة لغوية في العربية المعاصرة تقرب من مليون كلمة باستخدام تقنيات الإحصاء اللغوية؛ بإحصاء عن أكثر ١٠٠٠ كلمة تكرارًا في هذه المدونة، وأكثر ٥٠٠ متلازم دورانًا فيها، وأكثر ١٠٠ مسكوك استخدامًا فيها، وأكثر ١٠٠ أداة نحوية استخدامًا على مستوى هذه المدونة.

وبعد فقد قدم الكتاب رؤية مؤلفه لتقييم التقنيات الموجودة ووسائل تطويرها ومستقبلها، آملًا أن تكون لبنة في صرح كبير لمعالجة العربية وحوسبتها. والله ولي التوفيق.



# فهرس المراجع والمصادر

## فهرس المراجع

**أولاً: المراجع**

الإتقان في علوم القرآن، عبد الرحمن بن أبي بكر، جلال الدين السيوطي، تحقيق: محمد أبو الفضل إبراهيم، دار التراث، د.ط، د.ت.

استخدامات الحروف العربية (معجميا، صوتيا، صرفيا، نحويا،كتابيا)، سليمان فياض، دار المريخ، السعودية، د.ط، ١٩٩٨.

الإملاء والترقيم في الكتابة العربية، عبد العليم إبراهيم، مكتبة غريب، القاهرة، ١٩٧٥م.

البرهان في علوم القرآن، محمد بن بهادر بن عبد الله الزركشي، تحقيق: محمد أبو الفضل إبراهيم، دار المعرفة - بيروت، د.ط ، ١٣٩١هـ .

دليل الحيران شرح مورد الظمآن، إبراهيم بن أحمد المارغني التونسي، مكتبة الكليات الأزهرية - القاهرة - ١٩٨١م.

علم الدلالة، د. أحمد مختار عمر، عالم الكتب، القاهرة، ط٥، ١٩٩٨م.

معجم التعابير الاصطلاحية في العربية المعاصرة، د. وفاء كامل، القاهرة، ط١، ٢٠٠٧.

المعجم السياقى للتعبيرات الاصطلاحية، محمود إسماعيل صيني وآخرون، مكتبة لبنان ناشرون، بيروت، ط.١، ١٩٩٦م.

المعجم الوسيط، إبراهيم مصطفى، أحمد حسن الزيات، حامد عبدالقادر، محمد علي النجار، مجمع اللغة العربية بالقاهرة، ط.٤، ٢٠٠٤.
النحو الأساسي، د. أحمد مختار، د. مصطفى النحاس، د. محمد حماسة، دار السلاسل، الكويت، ١٩٩٤.
النحو الوافي. د. عباس حسن. دار المعارف. ط٣. مصر، القاهرة، ج٢. ص٤٦٦ ـ ج٣.

**ثانيا: الرسائل الجامعية**

اللبس في البنية والتركيب؛ أنماطه وأسبابه وآثاره وطرق التخلص منه، دراسة نحوية حاسوبية في العربية المعاصرة، عمرو جمعة عبد الرسول، رسالة دكتوراه بكلية دار العلوم، جامعة القاهرة، ٢٠١٤.
اللغة العربية المعاصرة في المجال السياسي، دراسة لغوية حاسوبية من واقع ذخيرة لغوية، هبة سيد راشد، رسالة ماجستير، كلية الألسن، جامعة عين شمس، ٢٠٠٨.
مدونة معجم عربي معاصر، معالجة لغوية حاسوبية، المعتز بالله السعيد طه، رسالة ماجستير، كلية دار العلوم، جامعة القاهرة، ١٤٢٨هـ/ ٢٠٠٧م.

**ثالثا: الأبحاث العلمية**

الأسس النظرية لوضع معجمٍ للمتلازمات اللفظية العربية، أسس المعجم النظري، د. محمد حلمي هليّل، مجلة المعجَمية، تونس، ١٩٩٧م، عــ ١٢، ١٣.
بحوث ومقالات في اللغة ، د. رمضان عبد التواب، مكتبة الخانجي بالقاهرة، ط.٣، ١٤١٥هـ -١٩٩٥م .
بعض صور التعبيرات الاصطلاحية في العربية المعاصرة، د. وفاء كامل، مجلة مجمع اللغة العربية بدمشق، المجلد ٧٨، الجزء٤.
تطوير التعرف الآلي على الحروف العربية، عمرو جمعة عبد الرسول، الندوة الدولية السادسة لعلوم وهندسة الحاسوب ، ٢٠١٠ تونس.
التعابير الاصطلاحية والسِّياقية ومعجم عربي لها، د. علي القاسمي ، بحث بمجلة اللسان العربي، مكتب تنسيق التعريب بالرباط، د.ت.
التقاء الحروف المتماثلة في القرآن الكريم، د.محمد زكي خضر ، مجلة الفرقان، العدد ٥٠، آذار ٢٠٠٦.
قواعد البيانات المحملة على الأقراص المدمجة: دراسة حالة لقاعدة بيانات الإنتاج الفكري الإسلامي INDEX ISLAMICUS ON CD-ROM ، هاشم فرحات، مجلة

مكتبة الملك فهد الوطنية، مجلد ٦، عدد ٢ (رجب ـ ذو الحجة ١٤٢١هـ/ أكتوبر ٢٠٠٠ـ مارس ٢٠٠١م).

المعايير اللغوية والفنية لتقييم برامج التعرف الضوئي على الحروف العربية، عمرو جمعة عبد الرسول، مؤتمر جمعية هندسة اللغة العاشر، ٢٠١٠ القاهرة.

من مشكلات التحليل النصي للمحتوى العربي على شبكة الإنترنت د.سلوى السيد حمادة، عمرو جمعة، مؤتمر هندسة اللغة ٢٠٠٨.

من مشكلات التحليل النصي للمحتوى العربي على شبكة الإنترنت، د.سلوى السيد حمادة، - عمرو جمعة، المؤتمر الثامن للجمعية المصرية لهندسة اللغة (كلية الهندسة- جامعة عين شمس) - ٢٠٠٨.

**رابعا: المواقع والصفحات على الإنترنت**

http://faculty.ksu.edu.sa/Alhendawy/My%20publications/%D9%82%D9%88%D8%A7%D8%B9%D8%AF%20%D8%A7%D9%84%D8%A8%D9%8A%D8%A7%D9%86%D8%A7%D8%AA.pdf

http://wiki.arabeyes.org/مستبعدات_الفهرسة

http://www.al-mishkat.com/khedher/wp-admin/Papers/paper18.pdf

http://www.coltec.net/default.aspx?tabid=224.

http://www.coltec.net/Portals/0/COLTEC_PDFs/ProofingTool_NEW.pdf

http://www.comp.leeds.ac.uk/eric/latifa/research.htm

http://www.concordancesoftware.co.uk/

http://www.cro-code.com/textanz.jsp

http://www.diwanalarab.com/IMG/pdf/Is_hamaatUolamaaAltaumieat1-1.pdf

http://www.sakhr.com/arabicresources.aspx.

# فهرس الكتاب

# تقنيات اللغة العربية الحاسوبية

## معايير التقييم ورؤى التطوير

يهدف هذا الكتاب إلى تقديم منهج لغوي لتقييم التقنيات اللغوية الحاسوبية التي تعالج اللغة العربية، من خلال عدد من المعايير اللغوية، فنظرة واحدة على بعض التقنيات العربية كافية لإدراك الهوة الواسعة بين مخرجات هذه التقنيات وبين ما يجب أن تكون عليه النتائج المرجوة، ومن ثم دعت الحاجة إلى وضع لبنة لغوية لمعايير تقييم هذه التقنيات، وستقتصر هذه المعايير على الجانب اللغوي فقط تاركة الجانب الفني للمتخصصين فيه

كما يهدف الكتاب إلى تقديم رؤى تطويرية جديدة لتطوير بعض التقنيات اللغوية العربية، يمكن من خلالها رفع كفاءة هذه التقنيات وتحسين جودة مخرجاتها، كما هو الحال مع تطبيقات التعرف الآلي على الحروف العربية، أو يمكن من خلالها إبراز الدور التعليمي لهذه التقنيات الحاسوبية، كما هو الحال مع المدققات الإملائية والنحوية؛ إذ يمتد دور هذه التقنيات اللغوية العربية إلى الإسهام في تعليم مستخدميها من العرب قواعد العربية كذلك، إضافةً إلى دورها الأساسي في معالجة العربية آليًّا.

د. عمرو جمعة

الدراسات ٤

هذه الطبعة
إهداء من المركز
ولا يسمح بنشرها ورقياً
أو تداولها تجاريا




ص.ب ١٢٥٠٠ الرياض ١١٤٧٣
هاتف:٠٠٩٦٦١١٢٥٨٧٢٦٨ - ٠٠٩٦٦١١٢٥٨١٠٨٢
البريد الإليكتروني: nashr@kaica.org.sa






9786039072065